CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 124, No. 3, Pt. 1 of 2 September 2008

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 24, No. 3, Pt. 2 of 2 September 2008

JASA EXPRESS LETTERS SPECIAL ISSUE

SHALLOW WATER ACOUSTICS EXPERIMENT 2006

(Continued)

CONTENTS—*Continued from preceding page*

 [DOI: 10.1121/1.2972156] Published Online 28 August 2008

Overview of Shallow Water 2006 *JASA EL* Special Issue Papers

James Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 02543
jlynch@whoi.edu

Dajun Tang
Applied Physics Laboratory, University of Washington, Seattle, Washington
djtang@apl.washington.edu

To improve basic understanding of shallow water acoustics and its dependence on oceanographic and geological environments, the Office of Naval Research (ONR) sponsored a multidisciplinary, multi-institution, multi-national experiment off the coast of New Jersey in the summer of 2006, designated "Shallow Water '06" (SW06). This large-scale experiment had three named components corresponding to the three major research and technology thrusts: in acoustics, LEAR (for Littoral Environmental Acoustics Research), in physical oceanography, NLIWI (the Non Linear Internal Waves Initiative), and in vehicles, AWACS (for Acoustic Wide Area Coverage for Surveillance). This special issue contains a collection of papers primarily dedicated to the initial results on underwater acoustics. Research results on oceanography and geology related to SW06 and providing the environmental underpinnings of the SW06 acoustics are not included in this special issue; they are planned for future publications.

The goal of SW06 was to study in detail the impact of various environmental processes on acoustics. The specific site for the experiment, centered on the 80 m isobath of the U.S. continental shelf off Atlantic City, New Jersey, was chosen for three reasons. First was the existence of a number of interesting shelf and slope features and processes, such as non-linear internal waves, the shelfbreak front, and complex and interesting seabed geology. Another reason this site was chosen is the existence of substantial historical data sets and knowledge obtained from previous experiments, particularly regarding the bottom and sub-bottom, which is critical but expensive to obtain. Finally, the easy accessibility of the site to the research vessels of a number of universities and oceanographic institutions reduced logistical costs.

SW06 was conducted over two months (mid-July to mid-September) in 2006, and involved seven ships, sixty-two moorings, aircraft overflights and satellite coverage, a fleet of ten oceanographic gliders, real-time data communication from a number of shipboard experiments, and dozens of principal investigators tending to their parts of this coordinated effort.

The core measurement suite of the SW06 experiment was an array of 62 acoustical and physical oceanographic moorings deployed in a "T" geometry, with the stem of the "T" stretching 30 km along shelf at the 80 m isobath, and the top of the "T" stretching 50 km across shelf from the 500 m isobath to the 60 m isobath. The top of the "T" is a central area where assets were concentrated to provide extensive environmental measurements. The reason for this "T" geometry was to study both along-shelf and across-shelf environmental effects on acoustics. Acoustically, SW06 encompassed frequencies from a few hundred Hertz to 20 kHz. This community has termed the band up to 1 kHz as low-frequency, and the band from 1 kHz to 20 kHz as mid-frequency. Thus, along-shelf versus across-shelf and low- versus mid-frequency became two of the major themes of the experiment. A more detailed description of the experiment's scope and layout can be found in Ref. 1.

Most of the nineteen papers in this special issue can be divided into two categories: (*i*) acoustic propagation and scattering and (*ii*) geoacoustic inversion. Under acoustic propagation and scattering, the papers address issues involving the impact of linear and non-linear internal waves on sound propagation, the spatial and temporal coherence and fluctuations of sound fields, and the sensitivity of propagation to various environmental parameters. The papers on geoacoustic inversion concentrate on determining an elasto-acoustic model of the bottom (shear and compressional wave speeds and attenuations, along with the density) based on vari-

ous types of acoustic data. Let us now look briefly at précis of the papers, by way of introduction. We will start with the propagation and scattering studies.

Luo *et al.* confirmed previous findings that non-linear internal waves focus and de-focus low-frequency sound that propagates parallel to the direction of the internal wave crests. This result is unique, in that the SW06 instrumentation suite provided a complete description of the internal waves from source to receiver, thus removing any experimental ambiguity from previously reported results. The papers by Rouseff *et al.* and by Katsnelson *et al.* deal with the unique scenario of mid-frequency acoustic propagation through a well characterized specific nonlinear-internal wave. The two papers by Tang *et al.* discuss the mean intensity and intensity fluctuations at fixed points over time due to the presence of linear, ambient internal waves, where nearly-saturated intensity fluctuation is found. Collis *et al.* observed horizontal field coherence and discuss low-frequency array performance in a time-varying internal wave field. DeFerrari *et al.*'s paper presents detailed observations of low-frequency temporal and spatial coherence. Taken together, these papers demonstrate both the diverse impact of internal waves on sound propagation and the progress in dealing with such problems both in measurements and modeling. The last environmentally focused paper is the one by Pecknold *et al.* which studies the sensitivity of longer range shallow water sound propagation to various environmental parameters.

Eight papers in this issue are devoted to geoacoustic measurements and inversions. A subgroup of these papers studies a localized area where multiple investigators performed mid-frequency acoustic measurements and carried out independent inversions on the same patch of seafloor, providing an opportunity for cross comparison of results. Yang *et al.* present *in situ* measurements of sediment sound speed at this central area to a depth of 1.7 m, using novel instrumentation. Turgut and Yamamoto present complementary *in situ* measurements at shallower sediment depths in the same area. Their paper concentrates on the investigation of the dispersion of sediment sound speed and attenuation in the frequency band of 10–80 kHz. The two papers by Choi *et al.* and Jiang *et al.* present inversions using mid-frequency sound where the acoustic sources and receiving arrays are separated by only 50 m to 500 m, offering a wide variety of incident angles of the sound on the bottom. The results of these four papers show consistent values of sediment sound speed, hence paving the way for future work on making direct comparative studies between *in situ* measurements and acoustic inversions.

Four papers in the bottom inversion group deal with low-frequency, longer range (order 10 kilometers) inversions. Ballard and Becker present a variant of the modal inversion techniques where one first estimates the normal mode eigenvalues (horizontal wavenumbers), and then uses these as data to invert for sediment sound speed as a function of depth. Their work in the present issue is for the more basic range independent case, with the aim of transitioning this work into the realm of range dependence. Potty *et al.* used a combustion source to generate low-frequency sound to invert for sediment sound speed using the mode dispersion curves as data, and studied the range-dependent medium case using adiabatic modes. The two papers by Knobles *et al.* are both concerned with sediment inversion over a sandy ridge track. One paper uses controlled sources and estimates sediment sound speed dispersion and attenuation, whereas the other uses ambient noise to investigate sediment attenuation frequency dependence.

The remaining three papers cover slightly different subjects. The paper by Dahl *et al.* is on field measurements and modeling of attenuation from near-surface bubbles at mid-frequencies, where excess attenuation from bubbles is seen. While sea surface related problems are not a large area of concentration for this special issue, surface scatter is an integral part of the shallow water environment that influences shallow acoustic propagation and reverberation, and significant resources were devoted to sea-surface characterization in SW06. The Traer *et al.* paper studies shallow-water seismo-acoustic noise generated by storms by using passive acoustics and correlating the ambient noise with wind storm measurements. Morozov *et al.*'s paper deals with underwater communications issues using modal processing. The latter two papers demonstrate some of the numerous applications of underwater sound. In both cases, knowledge of the

environment is crucial for success of the acoustic system, thereby reinforcing the importance of making adequate environmental measurements.

To summarize, this special issue describes recent progress made in understanding shallow water acoustics in a very complex coastal environment. Other significant results from this program, such as a multitude of three-dimensional acoustic effects caused by internal waves, range-dependent geoacoustic inversions, and backscatter modeling using measured sediment roughness are planned for future publications. The papers in this issue are very much the "first fruits" of the program.

Looking ahead, some of the immediate challenges in shallow water acoustics are (1) understanding and modeling sound propagation and its uncertainty in the presence or absence of non-linear internal waves, (2) how to systematically deal with the three-dimensional effects of sound propagation caused by the environment, and (3) how to make statistically meaningful comparison between geoacoustic inversion and *in situ* measurements. The primary long-term challenge is how to determine and then make the necessary environmental measurements required to predict shallow water propagation and, especially reverberation, which encompasses all aspects of shallow water acoustics and requires finer environmental knowledge. Some substantial progress has emerged from the SW06 data and measurements, and we expect more progress will follow from experiments yet to come.

Acknowledgments

First, we would like to thank our ONR sponsors, particularly Drs. Ellen Livingston, Theresa Paluszkiewicz, and Tom Curtin for their support of the SW06 experiment, and their help in coordinating its efforts.

Next, the success of this very large-scale field project benefited enormously from the collaboration of all participants and shared equipment and resources. We thus thank all our SW06 collaborators, both Principal Investigators and support personnel, for their collegiality and generosity in the performance of this experiment.

Finally, we would like to thank Drs. William Carey, Keith Wilson, and Allan Pierce of JASA EL and JASA for their efforts to make this special issue a reality.

References and links

[1]D. Tang, J. Moum, J. Lynch, P. Abbot, R. Chapman, P. Dahl, T. Duda, G. Gawarkiewicz, S. Glenn, J. Goff, H. Graber, J. Kemp, A. Maffei, J. Nash, and A. Newhall, "Shallow Water '06, A joint acoustic propagation/ Nonlinear internal wave physics experiment," Oceanogr. **20**, No. 4, December 2007.

Observation of sound focusing and defocusing due to propagating nonlinear internal waves

J. Luo, M. Badiey, and E. A. Karjadi
College of Marine and Earth Studies, University of Delaware, Newark, Delaware 19716
luojing@udel.edu, badiey@udel.edu, karjadi@udel.edu

B. Katsnelson and A. Tskhoidze
Voronezh State University, Universitetskaya Sq. 1, Voronezh 394006, Russia
katz@phys.vsu.ru, tskhoidze@phys.vsu.ru

J. F. Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
jlynch@whoi.edu

J. N. Moum
College of Oceanic & Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331-5503
moum@coas.oregonstate.edu

Abstract: Fluctuations of the low frequency sound field in the presence of an internal solitary wave packet during the Shallow Water '06 experiment are analyzed. Acoustic, environmental, and on-board ship radar image data were collected simultaneously before, during, and after a strong internal solitary wave packet passed through the acoustic track. Preliminary analysis of the acoustic wave temporal intensity fluctuations agrees with previously observed phenomena and the existing theory of the horizontal refraction mechanism, which causes focusing and defocusing when the acoustic track is nearly parallel to the front of the internal waves [J. Acoust. Soc. Am., **122**(2), pp. 747–760 (2007)].

PACS numbers: 43.30.Re, 43.30.Es, 43.30.Dr, 43.30.Bp [WC]
Date Received: March 15, 2008 **Date Accepted:** June 4, 2008

1. Introduction

In shallow water regions, nonlinear internal solitary waves (ISWs) create an anisotropic water column that affects acoustic wave propagation. In recent years, experimental and modeling efforts[1–3] have been conducted to quantify acoustic wave interaction with ISWs. These studies suggest that there are three different mechanisms that can explain the intensity variations in the presence of ISWs. The adiabatic regime corresponds to fluctuations of the sound field due to local variations of the water column (i.e., sound speed and/or depth) at the position of the source or the receiver. Next, propagation can be dominated by the mode coupling regime, where the interaction with the internal waves can cause energy transfer between different modes which in turn can result in fluctuations in the intensity. Finally, there is the horizontal refraction regime in which the distribution of the acoustic field in the horizontal plane can cause fluctuations in the measured field. The last two situations result in more significant signal intensity fluctuations than the first.[1–3] While the comparison of modeling results with data have previously shown the validity of this theory, a direct simultaneous measurement of the acoustic field and the ISW has not yet been reported to establish the transition among different mechanisms.

In our previous work, we separated propagation regimes depending on the angle between the ISW wave front and the direction of the acoustic track. In the Shallow Water '06 (SW06) measurements studied in this paper, we present a data set where the acoustic track angle is kept unchanged while the internal wave slowly transects the acoustic track for the source-receiver geometry. When the ISW train is starting to cover the receiver array only and does not

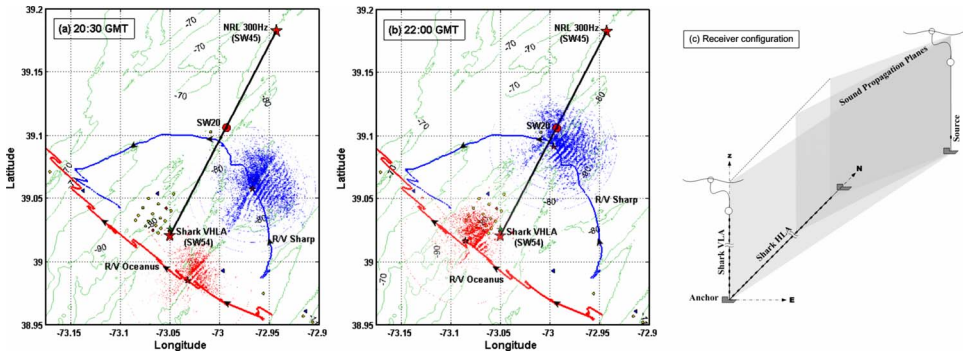


Fig. 1. (Color online) Map of the experiment location showing the positions of the acoustic source (NRL 300 Hz on SW45 mooring), the receiver array (Shark VHLA on SW54 mooring) and a midpoint thermistor array (SW20) together with the two research vessels (R/V SHARP and R/V OCEANUS) at two different geotimes. Times are (a) 20:30 GMT when the ISW has not reached the acoustic track, (b) 22:00:00 GMT when the ISW has occupied a large portion of the acoustic track including the receiver array. Ship tracks and radar images of R/V SHARP and R/V OCEANUS are shown. The configuration of the receiver Shark VLA and Shark HLA is shown in (c).

occupy all (or a significant part of) the acoustic track, we have the adiabatic regime. As the train moves and continues to cover the acoustic track, propagation changes to the horizontal refraction regime. Thus, we can isolate the transition between the two regimes.

The multi-institutional SW06 experiment was conducted on the New Jersey continental shelf from July to September 2006 at a location where internal wave activity has been observed and studied in the past.[4] In this paper, we focus on a particular internal wave event on August 17, 2006, for which a complete set of acoustic and environmental data were collected simultaneously. In addition, ISW surface signatures were captured continuously by the onboard radars of two research vessels prior to the arrival of, and during the passing of, the ISW packet over the acoustic track. The acoustic wave field variation is studied during this process. Other studies of the SW06 data in relation to three dimensional modeling are being conducted separately.[5]

2. Simultaneous acoustic and internal wave measurement

During SW06, both acoustic and environment data were collected simultaneously. In this paper we discuss the acoustic data from a particular acoustic source (NRL 300 Hz) on the mooring denoted SW45 (see Fig. 1). The source was located 72 m below the sea surface and 10.5 m above the sea floor at 39°10.957′ N, 72°56.575′ W, and transmitted 2.0489 second Linear Frequency Modulating (LFM) signals from 270 to 330 Hz every 4 s. Transmissions continued for 7.5 min and then repeated every half hour. A vertical and horizontal receiver array (the "Shark VHLA") on mooring SW54 was located at 39°01.252′ N, 73°02.983′ W, about 20.2 km south of the NRL source. The vertical part of the receiver array consisted of 16 hydrophones with 3.5 m spacing and was positioned in the water column from 13.5 to 77.75 m below the surface. The horizontal part of the array consisted of 32 hydrophones on the seafloor with spacing of 15 m, providing 478 m of horizontal aperture. The sampling rate of the array was 9765.625 Hz. The acoustic track and the locations of the source and receiver as well as the horizontal and vertical array configurations are shown in Fig. 1. The water depth along the acoustic track was about 80 m.

On August 17, 2006, at about 18:00 GMT, the R/V SHARP from the University of Delaware was located at 38°59.262′ N, 72°57.492′ W and the R/V OCEANUS from the Woods Hole Oceanographic Institution (WHOI) was located at 38°57.426′ N, 72°56.676′ W. Both platforms observed the origination of an ISW near the shelf break. This event was named Event 50 on R/V SHARP and ROSEY on R/V OCEANUS.

Figures 1(a) and 1(b) show the track of each vessel following this event from 18:00 GMT on August 17 to 02:00 GMT on August 18, 2006. The R/V SHARP's track was semicircu-

lar, centered at the WHOI vertical and horizontal line array (Shark VHLA) with the ship being positioned on the trough of the leading ISW front and moving with the advancing front. The R/V OCEANUS followed the same ISW packet from its initial location, while keeping a watch on the advancing wave front. Reversals in the track of R/V OCEANUS indicate where the wave packet was crossed, and during these periods, intensive profiling of temperature, density, turbulence, and sound velocity was conducted.

The two ships observed different parts of the ISW front, thus providing a large spatial coverage. The surface signatures of the ISW were digitally recorded by on-board ship radar images every 30 s. Combined radar images from the two vessels, each about 11.1 km in diameter, covered the receiver and about two thirds of the acoustic track. Mm. 1 shows a movie of the ships' radar images following the ISW during this event, from 18:00 GMT on 17 August to 02:00 GMT on 18 August, 2006. In this paper we discuss two situations: (1) when the ISW packet had not reached the acoustic track [time period T_{g1} (20:30 to 20:37 GMT)] and (2) when the ISW occupied most of the acoustic track [time period T_{g2} (22:00 to 22:07 GMT)]. The radar images from both vessels at 20:30 GMT and 22:00 GMT are shown in Figs. 1(a) and 1(b), respectively.

Mm. 1. Ships' radar movie during Event 50 (ROSEY) showing the movement of the internal wave. This is a file of type "gif" (10 MB).

At 20:30 GMT, the R/V SHARP radar showed ten distinct ISW fronts in a packet. At the same time, the R/V OCEANUS radar showed seven wave fronts with less curvature. The spacing between wave fronts in the packet varied from ~0.4 to 0.5 km for the leading wave fronts to ~0.2–0.3 km for the trailing waves. The ISW direction of propagation was about compass direction 300°.

At 22:00 GMT, most of the acoustic track including the receiver array was occupied by the ISW packet [see Fig. 1(b)]. The physical characteristics of the ISW packet during this time were similar to those at 20:30 GMT, except that its propagation direction was slightly changed. The R/V OCEANUS observed the ISW propagation direction close to the receiver to be about 310°. At the middle of the acoustic track (SW20), a propagation direction of 305° was observed on R/V SHARP. Note that the wave front was curved so there is a natural discrepancy between the propagation directions observed by the two ships. The angle between the ISW fronts and the acoustic track at midpoint was approximately 5° and varied along the acoustic track. The average speed of the ships following the ISW packet was about 1.8–2.0 knots, which was similar to the ISW propagation speed.

The SW45 mooring had 11 temperature sensors located at depth between 15 and 72 m and the SW54 mooring had 11 temperature sensors at 5–78.5 m. To get a spatial picture of the water temperature along the acoustic track, data from a third thermistor string (SW20) in the middle of the track was also used. This mooring had three temperature sensors between 14 and 40 m. All these sensors recorded temperature data every 30 s.

Figure 2 shows the temperature records at the acoustic source mooring SW45, the midpoint SW20, and the receiver mooring SW54 from 20:00 to 23:00 GMT on August 17, 2006. A solitary internal wave arrived at the receiver around 21:15 GMT, followed by a calm period of about 20 min. At around 21:40 GMT, the leading front of a strong ISW packet passed the receiver position. The ISW packet caused a sudden increase in the thermocline depth. At 22:20 GMT, the water temperature decreased slowly until 09:00 GMT the next day. The leading wave front was observed at SW20 at 22:02 and at SW45 at 22:15 GMT, respectively. If we assume the leading wave front is linear (it is not) and the wave speed is constant along the wave front, this suggests an angle of about 5° between the advancing front and the acoustic track.

Next, we show the acoustic signal arrival on the Shark VHLA during the two periods T_{g_1} and T_{g_2}. In order to show the intensity variations, we show the arrival of two different LFM pings on the array at geotimes separated by 102 s. In Fig. 3(a), the upper panel shows the acoustic signal on the VLA at 20:30:16 GMT while the lower panel shows the same signal arriving on the HLA. Figure 3(b) shows the same two arrays at 20:31:58 GMT. The parallel lines on the HLA data shown in the lower panels are due to the difference of arrival time on the HLA [see

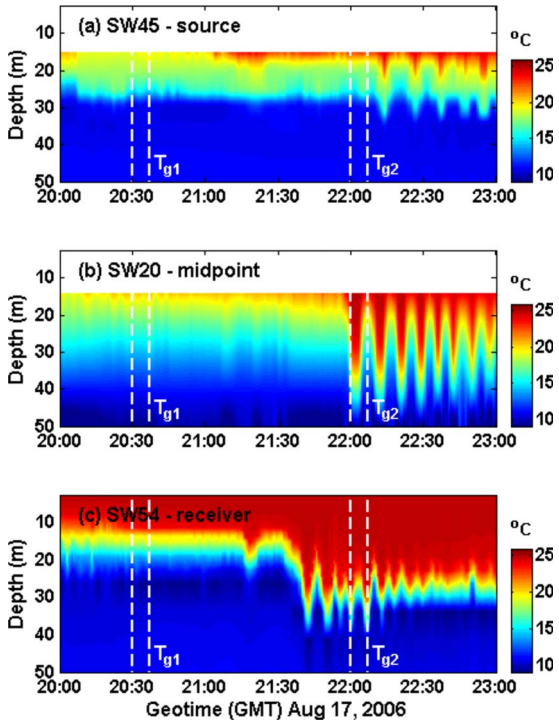


Fig. 2. (Color online) Temperature profiles measured at (a) the acoustic source, (b) the midpoint between the source and receiver, and (c) at the Shark WVHLA during internal wave Event 50, August 17, 2006 from 20:00 to 23:00 GMT. T_{g1}=20:30–20:37 GMT and T_{g2}=22:00–22:07 GMT.

array configuration in Fig. 1(c)]. The data on the VLA portion of the array shown in the upper panels best represent the modal arrivals. Note the similarity of the intensity values on both vertical and the horizontal array plots in Figs. 3(a) and 3(b), indicating the stability of the arriv-

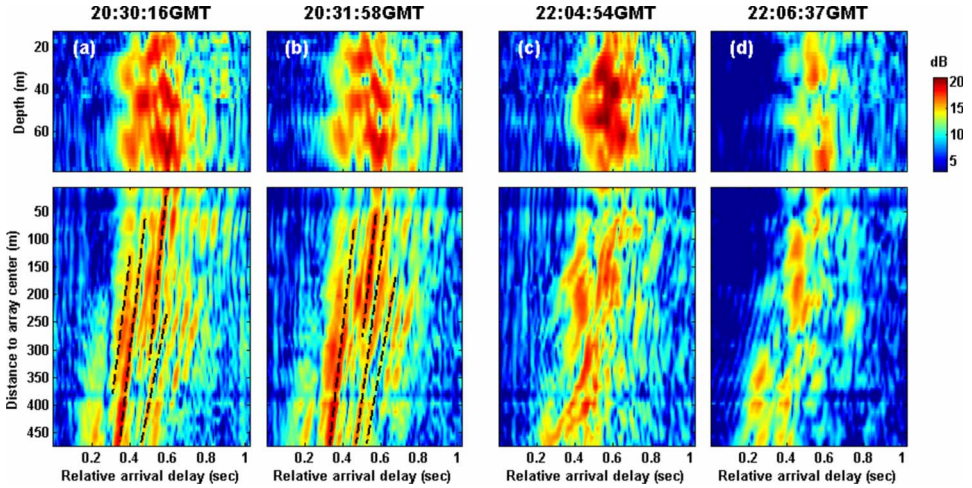


Fig. 3. (Color online) Received acoustic intensity on the Shark VHLA during two geotimes within the T_{g1} and T_{g2} periods. Upper panel shows the acoustic field on the VLA portion of the array while the lower panel shows the HLA portion. (a) T_g=20:30:16 GMT, (b) T_g=20:31:58 GMT, (c) T_g=22:04:54 GMT, (d) T_g=22:06:37 GMT.

ing signal energy during T_{g_1}. During T_{g_2} the same data format is shown. At 22:04:54 GMT shown in Fig. 3(c), both the VLA and HLA show a drastic change in the arrival. The signal intensity becomes very strong; however, different arrivals are mixed together and difficult to separate. The parallel lines on the HLA were notably distorted. At 22:06:37 GMT shown in Fig. 3(d), the arrivals show a lack of structure as well as lower intensity. During the time period of T_{g_2} the ISW is occupying a large fraction of the acoustic track. Mm. 2 shows a movie of the acoustic signal arrival on the Shark VHLA during the two periods T_{g_1} and T_{g_2}. To further quantify these results, we next calculate the average intensity in geotime and depth for the periods, T_{g_1} and T_{g_2}.

Mm. 2. Movie of the acoustic signal arrival on the Shark VHLA during the two periods T_{g_1} and T_{g_2}. This is a file of type gif (14 MB).

3. Intensity variations

Based on the theory proposed in our previous work,[3] the angle between the ISW fronts and the acoustic track determines the mechanism of the intensity fluctuations. Small angles provide horizontal refraction and focusing while larger angles cause mode coupling. Between these limits, there is an angular region for which the propagation is adiabatic. During internal wave Event 50, the ISW fronts passed through the acoustic track at an angle of about 5° providing the condition for horizontal refraction. Prior to the ISW arrival, we observe stable adiabatic propagation in a stable water column. Hence, a transition can indeed be observed between these two mechanisms.

To show a focusing event, we consider two geotimes: (a) T_{g_1} when the leading internal wave front had not reached the acoustic track and, (b) T_{g_2} when the internal wave packet occupied most of the acoustic track. We calculate the total intensity integrated over the depth H as

$$I(T) = \int_0^H I(z,T)dz, \tag{1}$$

where $I(z,T)=(1/\rho c)\int_{\tau}^{\tau+\Delta\tau} p^2(z,T,t)dt$ is the intensity of the signal arrivals integrated over a pulse length $\Delta\tau$ at a given depth z, where p is acoustic pressure, ρ is water density, and c is sound speed. The distribution of the ISWs in the horizontal plane can be assessed by carefully examining the radar image and the depth distribution of the temperature at three points (SW45, SW20, SW54) along the acoustic track.

The temperature distribution starting at 20:30:00 GMT (T_{g_1}) for the transmission period (i.e., ~7.5 min) at three different moorings shows very little fluctuation along the acoustic track (see Fig. 2). This indicates that the ISW has not reached the acoustic track during T_{g_1}, as is also shown in the radar image. In Fig. 4(a) we show a plot of $I(z,T)$ for the VLA. It is shown that for a repeated pulse of the same radiated intensity, only minor temporal intensity variation exists. In Fig. 4(b) we plot the depth integrated intensity for all geotimes, $I(T)$. The very small fluctuations (~3 dB) at T_{g_1} indicate a quiescent condition without ISWs in the track. This means that during this period of observation there is no redistribution of sound energy in the horizontal plane, i.e., there is no horizontal refraction. Small variations of the depth distribution of the sound intensity correspond to the adiabatic case. Figure 4(c) shows the acoustic intensity on the HLA. There are no apparent temporal intensity variations on the HLA during this geotime.

$I(z,T)$ for T_{g_2} is plotted in Fig. 4(d). Here we see increasing and decreasing trends in sound intensity over the VLA, which are synchronous in depth. The average intensity $I(T)$ using Eq. (1) peaks to ~35 dB around 22:04:30–22:05:00 GMT, and decreases to ~20 dB around 22:06:37 GMT as shown in Fig. 4(e). This significant fluctuation corresponds to redistribution of the acoustic energy in the horizontal plane, which in the limit can be referred to as focusing or defocusing events and is related to the position of the source and/or receiver with respect to

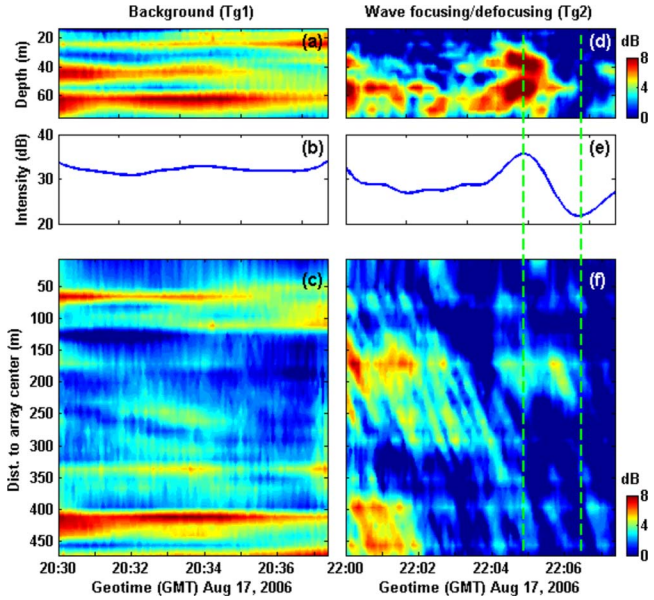


Fig. 4. (Color online) Received acoustic intensity during a 440 s (~7.33 min) transmission period for two geotimes, T_{g1} (from 20:30:00 to 20:37:20 GMT): (a) Depth distribution of total intensity per pulse $I(z,T)$, (b) depth integrated intensity, $I(T)$ for VLA, (c) total intensity $I(x,T)$ for HLA; and T_{g2} (22:00:00–22:07:20 GMT): (d) depth distribution of total intensity per pulse $I(z,T)$, (e) depth integrated intensity, $I(T)$ for VLA, (f) total intensity $I(x,T)$ for HLA.

the internal wave crests. The focusing and defocusing are also shown in Figs. 3(c) and 3(d). In this case (i.e., T_{g_2}), the receiver is between two adjacent maxima of the thermocline displacement, and the high intensity fluctuations (~15 dB peak to peak) are due to horizontal refraction effects similar to those shown in previous studies.[3] Temporal fluctuations correlate with oscillations of the thermocline layer at the receiver (see Fig. 2) due to ISWs (period is ~7–8 min). The fluctuations in the presence of internal waves (i.e., at T_{g_2}) are about 15 dB, which are much larger than those of T_{g_1} (~3 dB) with no internal wave in the acoustic track. Although the total jump in the thermocline thickness (i.e., from ~15 m to ~30 m) could make a difference in the "absolute" value of the acoustic intensity, it does not play a significant role in the temporal intensity fluctuation with periods corresponding to the ISW.

As in the quiescent case shown above, Fig. 4(f) shows the acoustic intensity on the HLA for the active period (T_{g_2}). During the geotime of about 7.33 min, we see the variation of the sound field at the horizontal array. For the variation observed during the first 3 min, there are two large intensity maxima observed at the HLA. These become weaker and disappear during the second half of the time period (after 22:03 GMT). This behavior, in our opinion, is a manifestation of horizontal redistribution of the sound intensity, which we call focusing/defocusing phenomena. The sloped intensity modulations correspond to the motion of an interference pattern registered on the horizontal array, which can be a result from any motion of the water layer (such as an ISW packet). The velocity of this interference pattern is estimated as 1–1.5 m/s from the acoustic data [shown as a slope in Fig. 4(f)]. This may not be directly related to the velocity of the internal solitons (measured from shipboard radar to be about 1 m/s). Further detailed modeling is needed to understand the nature of these fluctuations and their relationship to the slope of the lines shown in Fig. 4(f).

4. Summary

The SW06 experiment has provided high quality acoustic and environmental data to investigate the acoustic wave interaction with internal waves. Preliminary analysis of the acoustic data and

observations of radar images and temperature records shows that during the passage of an ISW event, horizontal refraction results in significant acoustic intensity variation. This observation agrees with the recently proposed theory of sound propagation through ISW.[3] Future work includes mode and frequency filtering of this acoustic data as well as modeling to establish the transition of acoustic field behavior from adiabatic to other mechanisms when an ISW passes an acoustic track.

Acknowledgments

The authors wish to thank all participants of the SW06 experiment, especially the scientific and ship personnel aboard the R/V SHARP and R/V OCEANUS. This research was supported by the Ocean Acoustic Program (321OA) of the Office of Naval Research through Grant No. N00014-07-1-0546.

References and links

[1]M. Badiey, Y. Mu, J. Lynch, J. R. Apel, and S. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**(1), 117–129 (2002).

[2]M. Badiey, B. G. Katsnelson, J. F. Lynch, S. Pereselkov, and W. L. Siegmann, "Measurement and modeling of three-dimensional sound intensity variations due to shallow-water internal waves," J. Acoust. Soc. Am. **117**(2), 613–625 (2005).

[3]M. Badiey, B. G. Katsnelson, J. F. Lynch, and S. Pereselkov, "Frequency dependence and intensity fluctuations due to shallow water internal waves," J. Acoust. Soc. Am. **122**(2), 747–760 (2007).

[4]A. E. Newhall, T. F. Duda, J. D. I. K. von der Heydt, J. N. Kemp, S. A. Lerner, S. P. Liberatore, Y.-T. Lin, J. F. Lynch, A. R. Maffei, A. K. Morozov, A. Shmelev, C. J. Sellers, and W. E. Witzell, "Acoustic and oceanographic observations and configuration information for the WHOI moorings for the SW06 experiment," WHOI technical report No. WHOI-2007-04, (2007).

[5]J. M. Collis, T. F. Duda, J. F. Lynch, and H. A. DeFerrari, "Observed limiting cases of horizontal field coherence and array performance in a time-varying internal wavefield," J. Acoust. Soc. Am. **124**(3), EL97–EL103 (2008).

Mid-frequency sound propagation through internal waves at short range with synoptic oceanographic observations

Daniel Rouseff, Dajun Tang, Kevin L. Williams, and Zhongkang Wang[a)]
Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698
rouseff@apl.washington.edu, djtang@apl.washington.edu, williams@apl.washington.edu; zkwang@apl.washington.edu

James N. Moum
College of Oceanic and Atmospheric Sciences, Oregon State University, 104 COAS Administration Building, Corvallis, Oregon 97331-5503
jmoum@coas.oregonstate.edu

Abstract: Preliminary results are presented from an analysis of mid-frequency acoustic transmission data collected at range 550 m during the Shallow Water 2006 Experiment. The acoustic data were collected on a vertical array immediately before, during, and after the passage of a nonlinear internal wave on 18 August, 2006. Using oceanographic data collected at a nearby location, a plane-wave model for the nonlinear internal wave's position as a function of time is developed. Experimental results show a new acoustic path is generated as the internal wave passes above the acoustic source.

PACS numbers: 43.30.Ft, 43.30.Re, 43.30.Xm [WC]
Date Received: February 23, 2008 **Date Accepted:** June 18, 2008

1. Introduction

In shallow water, there is extensive ongoing research into the impact of nonlinear internal waves on low-frequency (<1 kHz) sound propagation.[1–4] Issues that have been studied include acoustic mode coupling and horizontal refraction of ray paths. By comparison, less attention has been devoted to the effect of internal waves in the mid-frequency (1–10 Hz) band.

During the Shallow Water 2006 Experiment (SW06), mid-frequency acoustic transmission data were collected over a continuous 7 h period at range 550 m. The relatively short range was deemed desirable for isolating the effects of shallow water internal waves on acoustic propagation. At the SW06 site, both linear and nonlinear internal waves were potentially important. Linear internal waves often are modeled as a background random process introducing small changes in the sound speed that cause fluctuations in the acoustic field. At range 550 m, mid-frequency transmissions between 1 and 10 kHz were thought to span the transition between the regimes where classical weak- and strong-scattering theories for random media would apply.[5] Nonlinear internal waves are often modeled as a more event-like process causing strong, localized changes in the sound speed. Packets of nonlinear internal waves are not unusual and it was anticipated that a 550 m acoustic path might permit individual waves in the packet to be isolated.

This paper describes an initial analysis of acoustic data collected immediately before, during, and after the passage of a nonlinear internal wave. The results show that new acoustic paths are generated and that these new paths are particularly strong as the nonlinear internal wave passes above the acoustic source.

[a)]Permanent affiliation: Hangzhou Applied Acoustics Research Institute, 96 Huaxing Road, Hangzhou, China.

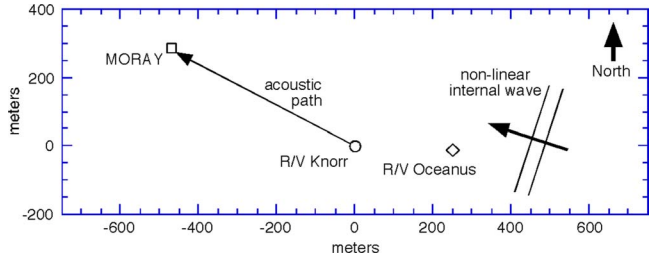


Fig. 1. (Color online) Experiment geometry. Acoustic source deployed off the stern of the R/V KNORR with transmitted signals measured 550 m away on MORAY vertical array. Oceanographic data were collected on R/V OCEANUS while a nonlinear internal wave passed.

2. Configuration of experiment

The SW06 experiment was performed in summer 2006 over the continental shelf of New Jersey, USA. The present analysis concentrates on data collected on 18 August in water nominally 80 m deep. When nonlinear internal waves were absent, the surface mixed layer extended to nominal depth 20 m.

The acoustic transmitter, an ITC-2015 (International Transducer Corporation) transducer, was positioned at 39 01.3152 N, 73 01.9364 W off the stern of the R/V KNORR. Deployment depth was 40 m to keep the transmitter below a thin layer of warm, salty water that was present that day at depth 30 m. Of present interest are linear frequency-modulated (LFM) chirp signals 20 ms in duration. The chirps spanned the 1.5–10.5 kHz frequency band with a raised cosine window and 10 percent taper. Transmissions were repeated approximately every 19 s.

The signals were recorded at range 550 m using the MORAY moored receiving system.[6] The system had two four-element vertical subarrays with the top four elements at depths 25.0, 25.2, 25.5, and 26.4 m, and the bottom four at 50.0, 50.2, 50.5, and 51.4 m. The data were stored at the array but a two-way radio frequency (rf) link permitted communication between MORAY and the R/V KNORR. The rf link was useful for adjusting amplifier gains, evaluating samples of the collected data, and assessing overall data quality.

While the acoustic data were being collected, investigators onboard the R/V OCEANUS collected oceanographic data (Fig. 1). At 20:53 Coordinated Universal Time (UTC), the R/V OCEANUS was positioned 250 m east of the R/V KNORR as a nonlinear internal wave first approached and then passed the two ships. The wave subsequently passed MORAY. X-band radar measurements estimated the bearing of the wave. Additional measurements included the water's temperature, salinity, density, and turbulence dissipation rate using the Chameleon turbulence profiler.[7] Two acoustic Doppler current profilers (ADCPs) were deployed to obtain vertical profiles of currents. A 120 kHz echosounder acoustically imaged the flow field.

3. Results

To understand the sequence of acoustic arrivals during quiescent periods when nonlinear internal waves were absent, the eigenrays from the source to different elements in the receiving array were calculated. The simulations assumed a range-independent environment and used three sound speed profiles collected from the R/V KNORR over a quiescent 4 h period. The simulations suggested that the direct acoustic path to the receiver at a depth of 50 m was the most sensitive to the details of the sound speed profile; depending on the particular profile, the direct path might or might not fragment into multiple direct arrivals. The fragmented direct path might arrive before or after the path that bounced off the sea surface. Furthermore, the direct path might be stronger or weaker than the surface bounce path. The acoustic ray that was launched at a downward angle and bounced once off the seabed, however, was strong and stable and insensitive to the particular sound speed profile used in the simulation.

Figure 2 shows a typical result from an acoustic transmission during a quiescent period. An LFM chirp signal measured at 50 m depth and 550 m range was matched filtered

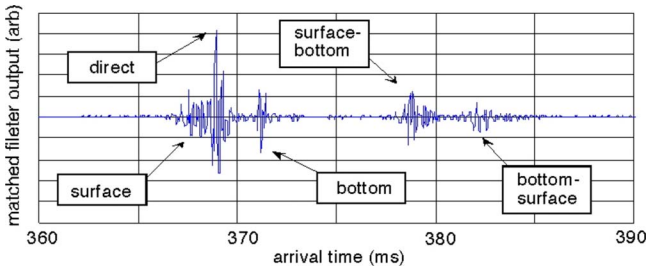


Fig. 2. (Color online) Typical acoustic arrival structure during quiescent period. Data collected on an array element at 50 m depth and 550 m range. Shown is the matched filter output where the various arrival paths are indicated.

yielding the arrival pattern for the different acoustic paths. The figure identifies the strongest direct path as well as four others: the sea surface bounce path, the bottom bounce path that reflected off the seabed, the surface-bottom bounce path and the bottom-surface path.

Figure 3 shows the fluid velocity measured from the R/V OCEANUS immediately before, during, and after the passage of nonlinear internal wave. Specifically, the vertical component of fluid velocity is mapped as a function of depth and time as measured by the ADCP. Isopycnals are superimposed. The maximum internal wave displacement is approximately 10 m and occurs at 21:14 UTC.

The oceanographic data collected on the R/V OCEANUS can be used to construct a simple model for the internal wave. When the internal wave was in the vicinity of the R/V OCEANUS, X-band radar measurements indicated the wave's bearing as 288 deg (Fig. 1) and its speed as 0.89 m/s. As an initial model, wavefront curvature is neglected and the internal wave is treated locally as a plane wave. The plane wave assumption, together with the known speed and direction of the wave, permits measurements made at the R/V OCEANUS to be propagated to other locations. Referring to the geometry in Fig. 1, the peak internal wave displacement observed on the R/V OCEANUS at 21:14 UTC will have propagated to the R/V KNORR at 21:18:30 UTC and to the MORAY at 21:28:31. The peak internal wave displacement, consequently, will lie somewhere between the acoustic transmitter and receiver for 10 min.

Figure 4 is a waterfall plot showing how the acoustic arrival structure measured at 50 m depth changes in time. The matched filter output is plotted for 100 consecutive LFM chirps with a gap of 18–20 s between transmissions. The 32 min of data include the periods immediately before, during, and after the passage of the internal wave. The superimposed horizontal lines bracket the time when the peak internal wave displacement, as calculated in the previous paragraph, is expected to pass between the acoustic source and receiver. The gross shifts in arrival time are due to relative motion between the source and the moored receiver. The bottom, bottom-surface, and surface-bottom bounce paths are labeled.

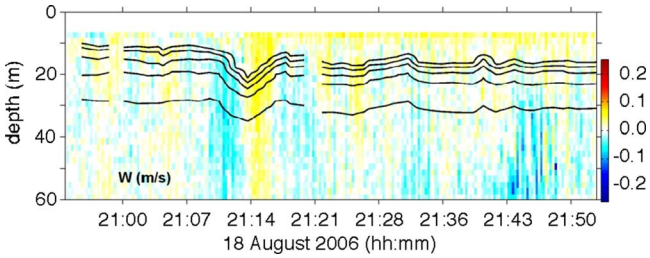


Fig. 3. (Color online) Sample oceanographic result showing vertical component of fluid velocity as measured on R/V OCEANUS. Contours of isopycnals are plotted over the color image. Peak isopycnal displacement from a nonlinear internal wave occurs at 21:14 UTC.

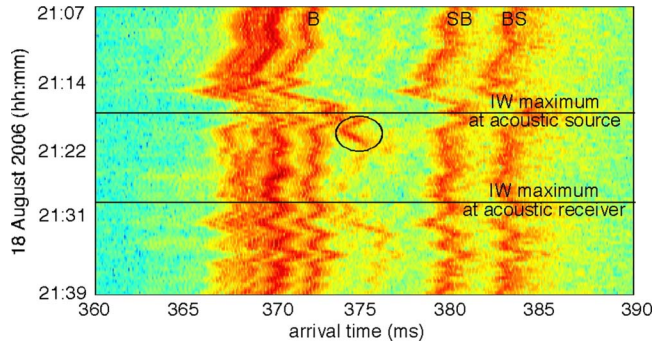


Fig. 4. (Color online) Time evolving acoustic arrival structure as the nonlinear internal wave in Fig. 3 enters the acoustic propagation path. The bottom, bottom-surface, and surface-bottom bounce paths are labeled. New acoustic path (circled) is generated as the internal wave passes above the acoustic source. Red-green-blue color scale has 50 dB dynamic range.

Similar arrival structures were observed on the other three elements in the bottom sub array. On each element, the bottom, bottom-surface, and surface-bottom bounce paths were strong and stable. Using the arrival times for a particular transmission as measured on each of the four array elements as input to a least-squares estimator, the arrival angle for a particular acoustic path could be estimated. Without correcting for array motion, the bottom bounce path had an average arrival angle of 8.4 deg relative to horizontal.

The most striking feature of Fig. 4 is the apparent generation of a new acoustic path as the internal wave passes above the acoustic source. The new path, circled in the figure as it appears, splits from the bottom bounce. At 21:21:13 UTC, approximately 3 min after the peak internal wave displacement has passed the acoustic source, the new path is at its strongest and its intensity exceeds the bottom bounce. Its arrival angle at 21:21:13 UTC is 12° and so steeper than the bottom-bounce path that precedes it. As the internal wave continues to move away from the acoustic source, the arrival angle for the new path continues to steepen but its intensity fades. It should be noted that a similar analysis was performed using the four elements in the upper sub array. On the upper sub array, whose shallowest element was at a depth of 25.0 m, the splitting of the bottom-bounce path was not observed as the internal wave passed.

A hypothesis is developed for the observed new ray. The passing internal wave depresses the mixed layer and hence the sound speed profile. Since the acoustic source is below the mixed layer, an upward-launched acoustic ray will be refracted downward by the passing internal wave. The refracted ray passes through an upper turning point and then strikes the bottom further down range than the original bottom-bounce path. After reflection, this new path arrives at the receiving array at a steeper angle than the unperturbed original bottom bounce.

4. Concluding remarks

Future work will include a more extensive data analysis together with numerical modeling. A complicating factor in developing a model is that the experiment was performed in a region where new individual internal waves were spawned rapidly.[8] Possible evidence for new internal waves is in Fig. 4 where a possible new acoustic arrival is apparent after the internal wave has passed the acoustic receiver at 21:28:31 UTC. Consequently, the simple plane-wave model for the internal waves used in the present work may prove inadequate for some calculations. To develop a more complete oceanographic model, temperature data collected on the MORAY acoustic array will be integrated into the analysis. A previously developed[9] ray tracing code valid for range-dependent media will then be applied in an attempt to reproduce and better understand the ray path generation observed in the data.

Acknowledgments

This work was supported by the Office of Naval Research.

References and links

[1]J. X. Zhou, X. Z. Zhang, and P. H. Rogers, "Resonant interaction of sound waves with internal solitons in the coastal zone," J. Acoust. Soc. Am. **90**, 2042–2054 (1991).
[2]R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, J. Apel, M. Badiey, C.-S. Chiu, S. Finette, M. Orr, B. Pasewark, A. Turgot, S. Wolf, and D. Tielbuerger, "Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220 (2000).
[3]M. Badiey, Y. Mu, J. Lynch, J. Apel, and S. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, 117–129 (2002).
[4]M. Badiey, B. G. Katsnelson, J. F. Lynch, S. Pereselkov, and W. L. Siegmann, "Measurement and modeling of three-dimensional sound intensity variations due to shallow-water internal waves," J. Acoust. Soc. Am. **117**, 613–625 (2005).
[5]B. J. Uscinski, *Elements of Wave Propagation in Random Media* (McGraw–Hill, New York, 1977), pp. 69–77.
[6]P. H. Dahl, J. W. Choi, N. J. Williams, and H. C. Graber, "Field measurements and modeling of attenuation from near-surface bubbles for frequencies 1–20 kHz," J. Acoust. Soc. Am. **124**, EL163–EL169 (2008).
[7]J. N. Moum, M. C. Gregg, R. C. Lien, and M. E. Carr, "Comparison of turbulence kinetic-energy dissipation rate estimates from two ocean microstructure profilers," J. Atmos. Ocean. Technol. **12**, 346–366 (1995).
[8]E. L. Shroyer, J. N. Moum, and J. D. Nash, "Observations of polarity reversal in shoaling non-linear internal waves," J. Phys. Oceanogr. , in press (2008).
[9]F. S. Henyey, D. Tang, K. L. Williams, R.-C. Lien, K. M. Becker, R. L. Culver, P. C. Gabel, J. E. Lyons, and T. C. Weber, "Effect of non-linear internal waves on mid-frequency acoustic propagation on the continental shelf," J. Acoust. Soc. Am. **119**, 3345 (2006).

Intensity fluctuations of midfrequency sound signals passing through moving nonlinear internal waves

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

James F. Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
jlynch@whoi.edu

Abstract: The fluctuations of intensity of broadband pulses in the midfrequency range (2–4.5 kHz) propagating in shallow water in the presence of intense internal waves moving approximately along the acoustic track are considered. These pulses were received by two separate single hydrophones placed at different distances from the source (~4 and ~12 km) and in different directions. It is shown that the frequency spectra of the fluctuations for these hydrophones have different predominating frequencies corresponding with the directions of the acoustic track. Comparisons of experimental results with theoretical estimates demonstrate good consistency.

PACS numbers: 43.30.Re, 43.30.Es, 43.30.Cq [DKW]
Date Received: March 26, 2008 **Date Accepted:** July 16, 2008

1. Introduction

Fluctuations of the intensity of low frequency sound waves in the presence of intense internal waves moving along an acoustic track were considered in previous papers[1] (experimental observations) and[2] (theoretical analysis). It was shown that the hypothesis of mode coupling as the physical reason for the fluctuations explains the peculiarities of the spectrum of these fluctuations (i.e., existence and value of a predominating frequency, the arrival times, etc.).

For mid and high frequency sound waves, ray language should be used for the description of sound propagation in the presence of internal solitons (IS) moving along the acoustic track[2] as illustrated in Fig. 1. A source and receiver (S and R) of sound signals are placed in shallow water with some distance between them, with trajectories of typical rays coming from the source to receiver as shown in Fig. 1. In the presence of internal solitons moving along the acoustic track, we have some moving perturbation of the thermocline layer and correspondingly some perturbation of the trajectory of the rays. This perturbation depends on the position of the IS and on the types of rays. For example, if the soliton is at position A, then it causes a distortion of ray 3, whereas for position B of the soliton, ray 3 does not interact with it. So during the motion of the soliton there are temporal fluctuations of the sound intensity at the receiver. It is clear that the perturbation depends on the type of rays; for example, ray 3 interacts with the soliton more strongly than rays 1 and 2. It has been shown that the most significant perturbation is provided by the rays which are tangent to the thermocline layer (these rays will be called as "critical rays," corresponding to ray 3 in Fig. 1), whereas other rays (denoted as 1 and 2 in Fig. 1) give a comparatively small contribution to the fluctuations. So if the phase velocity v_t of the IS crest along the direction of the acoustic track (which is greater than the real velocity v of the IS, which is perpendicular to the wave front of the IS) is taken to be constant, then the most significant fluctuations of the sound amplitude at the receiver will take place with a period T_c determined by v_t and the cycle distance of the critical ray D_c: $T_c = D_c / v_t$. For real ocean conditions, $T_c \geqslant 10$ min, and so during the time of observation, there can be 5–10 oscillations of the

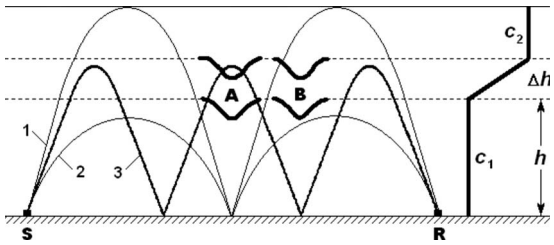


Fig. 1. Schematic of the interaction of an internal soliton with sound rays of different types radiated from a source placed at the bottom. Sound speed profile on the right side corresponds to a simplified one for SW06. We did not emphasize the small negative gradients in the upper and lower layers (seen in the experiment) in the profile graph, but they are evident in the ray trace results. Waveguide parameters are: $h \sim 60$ m, $\Delta h \sim 10$ m, $c_1 \sim 1480$ m/s, and $c_2 \sim 1525$ m/s.

sound intensity. A spectrum of this signal taken during the time of observation (~60 min) should contain set of maxima spanned by the value $\Omega_c \sim v_t / D_c$. The value of Ω_c is the "frequency of collisions" of the IS with the critical ray and can also be called the "predominating frequency." For typical conditions, $\Omega_c \leqslant 10$ cph.

An important feature of this spectrum is the dependence of the predominating frequency Ω_c on the angle between the direction of the acoustic track and the direction of propagation of the IS. Using this, it is possible to check this theory with an experiment where there are two acoustic tracks with different directions (two receivers for one and the same source) if they are crossed by the same soliton (or train of solitons). In this case we should have fluctuations of the same sequence of signals with different predominating frequencies, corresponding to the projections of the real velocity on the direction of the acoustic track. Such a situation took place in the SW06 experiment, along an acoustic track created by a source on the R/V KNORR.

2. The shallow water 2006 (SW06) experiment. Environmental data and soliton's parameters

The layout of the part of SW06 on 13 August during the time period 15:31–16:20 Greenwich mean time (GMT) is shown in Fig. 2. Positions of the source (R/V KNORR), two single hydrophone receiving units (SHRU1 and SHRU2) and thermistor strings (SW23, SW24, SW30,

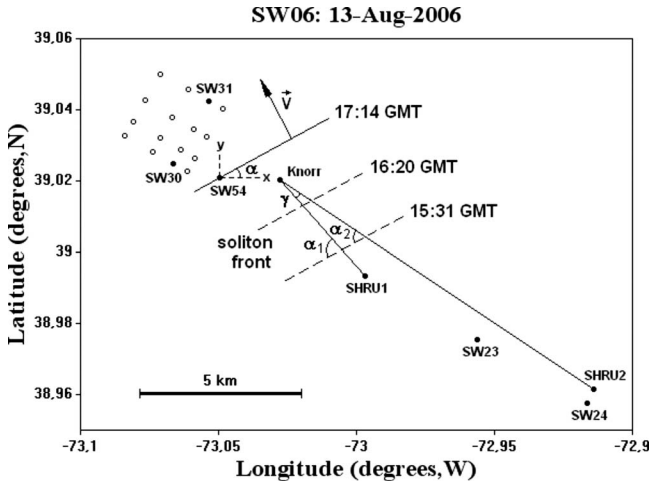


Fig. 2. Scheme of experiments with midfrequency pulses in SW06. Black straight lines denote acoustic tracks 1 (Knorr-SHRU1) and 2 (Knorr-SHRU2).

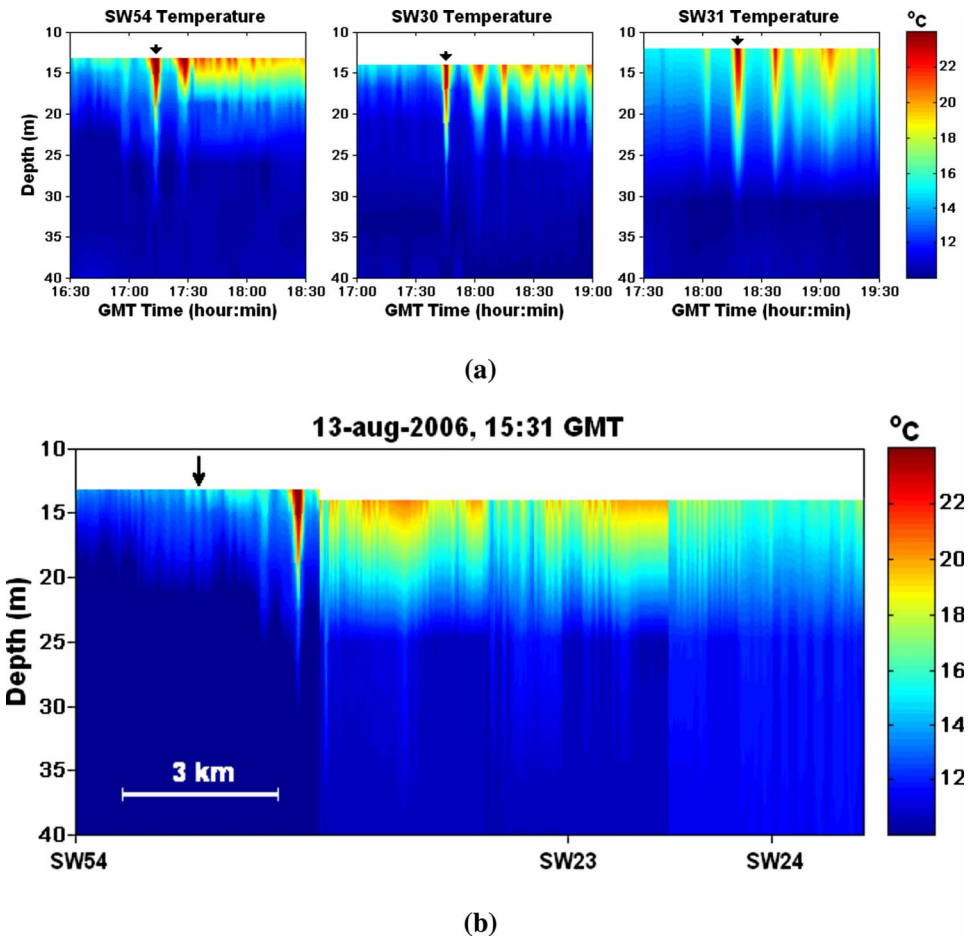


Fig. 3. (Color online) (a) Temperature records of an internal soliton at three thermistor strings. Color scale of temperature is shown at the right. (b) Reconstructed depth distribution of temperature for 15:31 GMT along a straight line between thermistor's sensors SW24-SW23-SW54. Black arrow denotes position of the soliton at 16:20 GMT.

SW31, SW54) are also shown in Fig. 2. The angle between the acoustic tracks is denoted by γ, and was found to be $\gamma \sim 15°$.

The behavior of internal waves was monitored by several thermistor strings, moored at different places. The temporal records of temperature depth distribution from the sensors denoted SW23, SW24, SW54, SW30, and SW31 (see Fig. 2) are used to find the shape of internal waves. During this time period, some solitons were registered. In Fig. 3(a) three temperature records of the soliton train are shown, which is moving approximately toward the coast. The moored sensors were reached by the first peak at different times [marked by arrows in Fig. 3(a)]. The different shapes of the temperature distribution demonstrate the evolution of the soliton train in time and space. It is easy to estimate the velocity and direction of propagation of the IS in a Cartesian coordinate system on the horizontal plane with the origin at mooring SW54. Suppose that the train or separate IS has a plane wave front directed at angle α with the X-axis, and moves with constant velocity v (Fig. 2). The given soliton consequently reaches sensor SW54, then SW30 and then SW31 [Fig. 3(a)]. The corresponding time differences between the points SW30 and SW31 from one side and SW54 from other side are $\Delta T_{30}=31$ min and $\Delta T_{31}=63$ min. In this case

$$\alpha = \arctan \frac{y_{30}\Delta T_{31} - y_{31}\Delta T_{30}}{x_{30}\Delta T_{31} - x_{31}\Delta T_{30}},$$

$$v = \frac{y_{30} - x_{30}\tan\alpha}{\Delta T_{30}} \cos\alpha,$$

where $x_{30}=-1477$ m, $y_{30}=456$ m, $x_{31}=-337$ m, and $y_{31}=2402$ m.

Taking into account experimental errors, we have approximately $\alpha \sim 29°$, $v \sim 0.6$ m/s.

For a more detailed analysis of internal waves outside of the acoustic track, the 14 thermistors SW04–SW17 can be used (shown in Fig. 2 by circles). In this case, the angle α changes during the motion from right to left in a sector from 26° to 44°, as a result of the variation of direction during the propagation and bending of the wave front. At the same time, along an acoustic track a comparatively constant direction of the wave front takes place, and for the following work we take $\alpha \sim 29°$.

Next, the spatial crossection of the internal waves along the straight line SW24-SW23-SW54 can be constructed for some fixed time, for example 15:31 GMT. Let us denote the temperature record at some point as $T^0(T,z)$, where T^0 is temperature and T is GMT. Within the framework of our approximation $T^0 = T^0[T - T_0 - (l - l_0)/v_l, z]$, where l is a length along the above mentioned direction, v_l is the corresponding projection of the real velocity v, and T_0, l_0 are some fixed GMT time and spatial position of the internal waves. Note that in spite of Fig. 3(a) showing some evolution of the IS propagating toward the coast, during our time of processing (~ 1 h) we can neglect this modification and consider the shape of the IS to be constant.

The reconstructed temperature distribution is shown in Fig. 3(b). According to our environmental picture, there is only one separate soliton at both acoustic tracks: KNORR-SHRU1 and KNORR-SHRU2 for the time interval, (15:31–16:20 GMT). Positions of the wave front of the soliton for 15:31 GMT and 16:20 GMT are shown in Fig. 2 by dotted lines and are also denoted in Fig. 3(b) (soliton itself and black arrow). The velocity of the soliton along this direction is about 0.7 m/s.

3. SW06 experiment. Midfrequency acoustical data and analysis

During our study period, a sequence of pulses was radiated consisting of 200 separate pulses in the frequency band 2–10 kHz. As was said above, these pulses were received independently at SHRU1 and SHRU2. The spectrum of radiated signals was in the band 2–10 kHz. However, for the processing, we use a sampling time $\delta t = 0.000\,102\,4$ s, which corresponds to maximal frequency in the spectrum of $(2\delta t)^{-1} = 4.88$ kHz. So only the intensity of received pulses in the band 2–4 kHz is considered here. The analysis above shows that the character of the fluctuations should not depend on the frequency in this band.

Let the received pulse have the sound pressure signature $p(t)$ with the spectrum

$$S(\omega) = \int_0^\tau p(t)\exp(i\omega t)dt. \tag{1}$$

Here, we take $\tau = 0.1$ s (the approximate duration of the pulse) for both receivers. Formally speaking, an infinite time period should be taken; however, it is more sensible to take this reduced period because outside of this interval there is noise only. So the energy of the pulse (up to constant factor) is

$$E = \int_{\omega_1}^{\omega_2} |S(\omega)|^2 d\omega, \tag{2}$$

where $\omega_1 = 2$ kHz and $\omega_2 = 4$ kHz.

The energy of the received pulses fluctuates from pulse to pulse, as a function of GMT, so it is denoted as $E=E(T)$. The temporal dependence of the relative fluctuations is

$$\bar{E}(T)=\frac{E(T)-\langle E\rangle}{\langle E\rangle}, \tag{3}$$

where $\langle E\rangle=1/\Delta T\int_0^{\Delta T}E(T)dT$ is the average energy for the time period $\Delta T=49$ min. The temporal dependence of the fluctuations of energy for both hydrophones is shown in Fig. 4

We see from Figs. 4(a) and 4(b) that the oscillations of the temporal dependence of the sound energy have a rather regular character. According to theory,[2] the frequency spectrum of the temporal behavior has predominating frequencies, corresponding to the quasiperiodical interaction of the sound pulse with the periodical ray structure of the sound field. The most significant interaction takes place for the critical ray, so the predominating frequency is determined by the time of motion of the pulse between adjacent maxima of the critical ray.

Let us consider the spectrum of fluctuations, or more exactly the amplitude of this spectrum

$$\bar{G}(\Omega)=\left|\int_0^{\Delta T}\bar{E}(T)\exp(i\Omega T)dT\right|. \tag{4}$$

The spectra for the temporal series from SHRU1 and SHRU 2 are shown in Fig. 4(c) in decibel (dB) units; more exactly, the values $10\log\bar{G}(\Omega)$ are shown.

It is seen that both spectra are functions having a quasiperiodical structure in the frequency domain in the region below 30 cph. The positions of the maxima in the spectra corresponding to SHRU1 and SHRU2 are shifted. According to interpretation,[2] this is due to the different directions of the sound propagation relative to the direction of propagation of the internal solitons.

The predominating frequencies and the average shifts of the spectra can be estimated from the experimental data [Fig. 4(c)]. In particular maxima (harmonics) number 4 are shifted by 1.7 cph. For this case, $\Omega_c^1\approx 3.58$ cph and $\Omega_c^2\approx 4$ cph.

Let us consider the temporal dependence from the point of view of our theory, i.e., that the motion of the soliton is the reason for the fluctuations with a predominating frequency, and compare the estimation from Ref. 2 with the results of our observations.

A diagram of the experiment showing all the relevant angles is found in Fig. 2. Here, $\gamma\sim 15°$ is the angle between acoustic tracks 1 and 2; $\alpha_{1,2}$ are the angles between tracks 1 or 2 and the wave front of the IS correspondingly; these angles can be found if we know angle α, so that $\alpha_1\sim 78°$ and $\alpha_2\sim 63°$.

The velocities of the IS along acoustic tracks 1 and 2 can be calculated using the $v_{1,2}=v/\sin\alpha_{1,2}$. The corresponding predominating frequencies of fluctuation for the two acoustic tracks are then $\Omega_c^{1,2}=v/(D_c\sin\alpha_{1,2})$, For the sound speed profile in the SW06 area, $D_c\sim 2hc_1/\sqrt{c_2^2-c_1^2}+2\Delta h\sqrt{c_2^2-c_1^2}/\Delta c$. For parameters in SW06 area $D_c\sim 640$ m. Calculations give for frequencies $\Omega_c^1\approx 3.45$ cph and $\Omega_c^2\approx 3.8$ cph. These values are very close to the measured ones. We remark that these values are sensitive to the value of the ray cycle distance. However, the ray cycle can be effectively excluded by considering only the ratio of the predominating frequencies

$$\Omega_c^1/\Omega_c^2=\sin\alpha_2/\sin\alpha_1. \tag{5}$$

For our case, this ratio is $\sin\alpha_2/\sin\alpha_1\sim 0.91$ and for the experimental data it is $\Omega_c^1/\Omega_c^2=0.895$, close to the aforementioned ratio.

4. Summary

In summary, we see that the IS, moving approximately along an acoustic track, interacts with the quasiperiodical spatial interference pattern of the sound field formed by a set of rays (in the

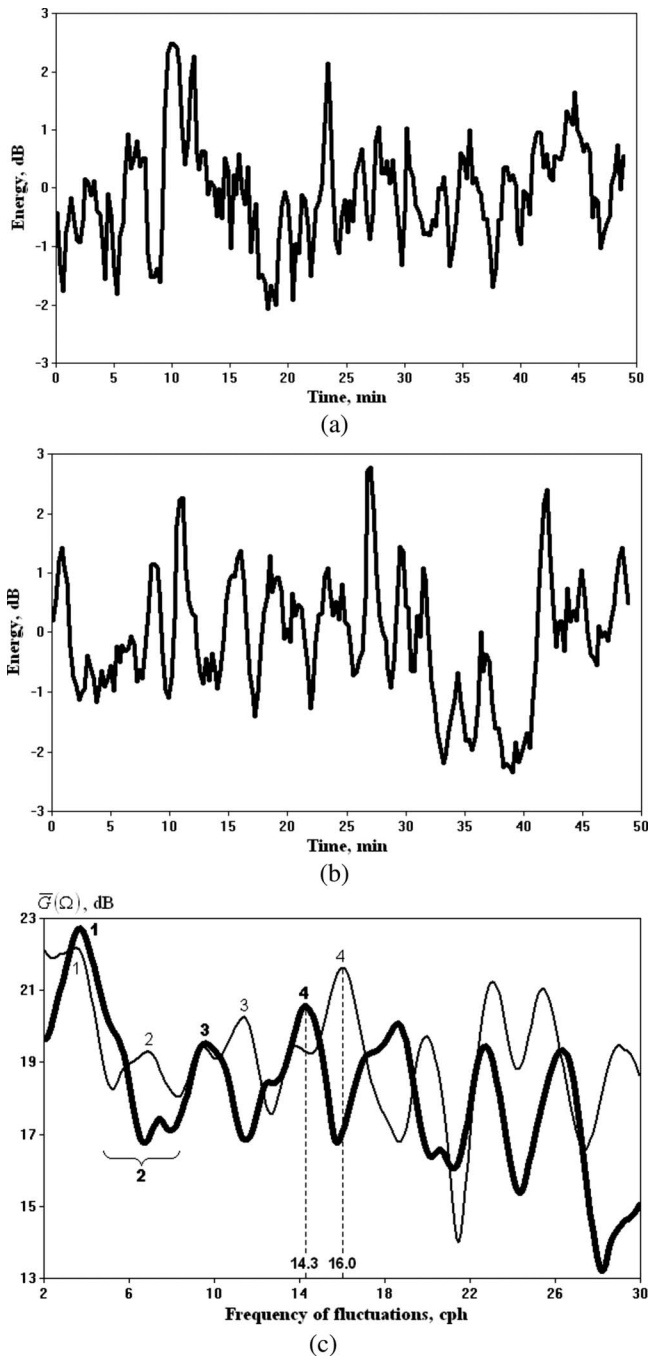


Fig. 4. Temporal dependence of the energy of the sound pulses for SHRU1 (a) and SHRU2 (b) for the time period 13 August 2006, 15:31–16:20 GMT. (c) Spectra of fluctuations of pulse energy. Bold line denotes spectrum for SHRU1, thin line corresponds to SHRU2. Dotted lines denote positions of the fourth harmonics of predominating frequencies.

high frequency approximation) radiated from the source in a shallow water waveguide. Correspondingly, the frequency of the temporal fluctuations of the sound intensity at the receiver is

proportional to the velocity of the IS along the acoustic track. The difference in spectra of these fluctuations (the shift of the predominating frequencies) for two acoustic tracks depends on the angle between them and the observations reported here are consistent with this hypothesis.

It is very interesting that the value of the predominating frequency does not depend (or depends weakly) upon the shape and amplitude of the IS, and on how many separate peaks there are in the train, and on its exact position between source and receiver. It is just this circumstance that allows us to get comparatively good agreement between theory and experiment.

We note also that this effect can also be used for solution of the inverse problem—to diagnose parameters of moving solitons (velocity and direction) on the basis of acoustical measurements with high (mid) frequency sources and two differently directed acoustic tracks.

Acknowledgments

The authors are very grateful to D. J. Tang for the information about the radiated signals and helpful discussions. This work was supported by the Office of Naval Research (ONR), the Russian Foundation for Basic Research (RFBR), Grant No. 06-05-64853, and CRDF (REC 010).

References and links

[1]M. Badiey, Y. Mu, J. Lynch, J. R. Apel, and S. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, 117–129 (2002).

[2]V. Grigorev and B. Katsnelson, "Fluctuations of high frequency acoustic signals in shallow water due to motion of internal soliton," Acoust. Phys. **54**, N6 (2008).

Mid-frequency acoustic propagation in shallow water on the New Jersey shelf: Mean intensity

Dajun Tang, Frank S. Henyey, Zhongkang Wang,[a] Kevin L. Williams, Daniel Rouseff, Peter H. Dahl, Jorge Quijano,[b] and Jee Woong Choi[c]

Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698
djtang@apl.washington.edu, frank@apl.washington.edu, williams@apl.washington.edu, rouseff@apl.washington.edu, dahl@apl.washington.edu

Abstract: Mid-frequency (1–10 kHz) sound propagation was measured at ranges 1–9 km in shallow water in order to investigate intensity statistics. Warm water near the bottom results in a sound speed minimum. Environmental measurements include sediment sound speed and water sound speed and density from a towed conductivity-temperature-depth chain. Ambient internal waves contribute to acoustic fluctuations. A simple model involving modes with random phases predicts the mean transmission loss to within a few dB. Quantitative ray theory fails due to near axial focusing. Fluctuations of the intensity field are dominated by water column variability.

PACS numbers: 43.30.Ft, 43.30.Re, 43.30.Bp, 43.30.Ma [WC]
Date Received: April 14, 2008 **Date Accepted:** May 21, 2008

1. Introduction

One of the major goals in ocean acoustics is to understand transmission loss (TL). This includes the mean intensity and the fluctuation of intensity. In shallow water, most work has been in the low-frequency band (<1 kHz).[1–3] Narrowband low-frequency intensity fluctuations due to water column variability have been investigated.[4,5] Mid-frequency (1–10 kHz) propagation has received much less attention either theoretically or experimentally. Ewart and Reynolds[6] measured specifically mid-frequency intensity fluctuations in the deep ocean, and quantified the intensity scintillation index.[7] In shallow waters, there is a dearth of measurement data in the mid-frequency range, especially those which measure intensity versus time at fixed ranges, necessary for statistical analysis.

During the Shallow Water '06 experiment (SW06), on August 11, 2006, a mid-frequency acoustic propagation experiment was performed in the frequency band of 2–10 kHz and at ranges 1–9 km. First, a 1-km propagation experiment, accompanied by detailed two-dimensional (2D) sound speed measurements from a towed CTD chain (conductivity, temperature, and depth) along the anticipated nonlinear internal wave travel direction was conducted. Then, propagation paths were chosen along the 80-m isobath for $1, 2, \ldots, 9$ km ranges. At each of these stations, 100 identical acoustic transmissions were made in order to study the mean and fluctuations of the intensity. This paper will provide details of the experiment setup and environmental measurements supporting the acoustics, and will also present the results on mean sound intensity over range and frequency. A companion paper will address the issue of intensity fluctuations.

2. Experiment description

At the experiment site, a typical strong summer thermocline existed, and in addition, warm, salty slope water underlay the colder mid-depth water. Thus, there was a sound speed minimum

[a]Permanent affiliation: Hangzhou Applied Acoustics Research Institute, 96 Huaxing Road, Hangzhou, China.
[b]Affiliation: Portland State University, 1900 SW Fourth Avenue, Suite 160-17, Portland, Oregan 97207.
[c]Current affiliation: Department of Environmental Marine Sciences, Hanyang University, Korea.

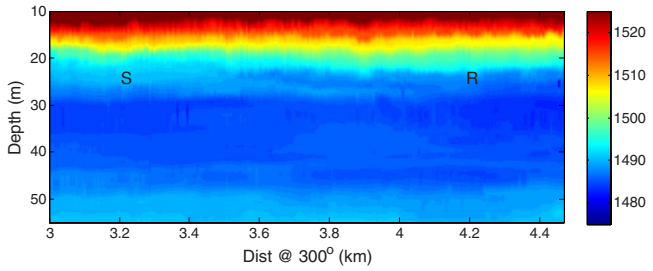


Fig. 1. (Color online) Sound speed profile measured by a towed CTD chain. The colorbar is in the units of m/s for sound speed. The horizontal axis is along the acoustic path, and defined as the projection on the 300° compass direction referenced to 39° N, 73° W as origin. The time of the measurement is 17:53–18:02 UTC on August 11, 2006. The letters "S" and "R" indicate the source and receiver positions for the first segment of the propagation experiment.

at ~32 m depth, forming a sound channel. This day was very unusual for SW06 because there were very few nonlinear internal waves. The combination of the sound channel and ambient internal waves had a large impact on mid-frequency sound propagation.

The mid-frequency sound source, an ITC-2015 transducer, was deployed off the stern of the R/V KNORR at 30 m depth, close to the sound axis to emphasize the propagation of energy trapped in the channel. The waveform transmitted consisted of two signals separated by 100 ms. The leading signal is a 10-ms multitone with nine frequencies, spanning 2–10 times the base frequency of 1011 Hz. The trailing signal is a 20-ms linear frequency-modulated chirp, which spanned the 1.5–10.5-kHz frequency band with a Hann window and 10% taper. Transmissions were repeated about every 20 s.

The signals were recorded on a moored receiving array system,[8] which was moored throughout the experiment at 39.0245 N, 73.0377 W. The system had two four-element vertical subarrays with the top four elements at depths 25.0, 25.2, 25.5, and 26.4 m, and the bottom four at 50.0, 50.2, 50.5, and 51.4 m. The data were stored at the array.

Both water column and bottom environmental data were collected to support modeling of the acoustics. Surfacial sediment sound speed was measured to be 1610–1620 m/s in the experiment area using a system called SAMS.[9] In addition to regular CTD casts, 2D sound speed profiles in the water column were measured by a towed CTD chain deployed from the R/V ENDEAVOR for a set of acoustic measurements at the 1-km range. An example of the 2D sound speed profile is given in Fig. 1, where the channel is clearly shown. In addition, it is evident that heterogeneity in sound speed exists as a function of range. The towed CTD chain data are available for several hours. After minor corrections, they will be used to simulate sound transmission through heterogeneities to statistically model the acoustic data.

The propagation experiment was divided into two segments. The first was at 1-km range over an extended period of time, lasting over 6 h and including 1500 ping repetitions. This provides ample data for statistical analysis of field fluctuations. It was during this segment that the accompanying towed CTD chain data were available. The second segment was at ten stations along the 80-m isobath at ranges $1, 2, \ldots, 10$ km. At each station, 100 pings were transmitted [except at the 10-km station (see Table 1), where a failure to communicate with the receiver limited the data to only eight pings].

3. Results

Ray tracing normally applies to medium frequencies. In Fig. 2, a fan of rays is calculated to a range of 2 km based on the range-independent sound speed profile shown on the right of the figure from a CTD cast. In this figure, the sound source is at 30 m depth, and rays span ±23°, corresponding to all rays with bottom grazing angle smaller than the critical angle (22°). For the receiver depth near the sound axis, shallow angle rays dominate. At about 1 km a strong focus can be seen near 30 m depth, and a similar one near 2 km, close to ranges used in the experi-

Table 1. Acoustic source locations and number of transmissions.

Station	Lat (N)	Lon (W)	Pings	Compass bearing
1 km	39.0203	73.0277	1500	300°
1 km	39.0325	73.0316	100	30°
2 km	39.0404	73.0259	100	30°
3 km	39.0482	73.0201	100	30°
4 km	39.0559	73.0143	100	30°
5 km	39.0637	73.0085	100	30°
6 km	39.0715	73.0028	100	30°
7 km	39.0776	72.9983	100	30°
8 km	39.0871	72.9912	100	30°
9 km	39.0948	72.9853	100	30°
10 km	39.1021	72.9794	8	30°

ment. The multiple caustics in the focusing regions imply that that mid-frequency sound intensity levels cannot be accurately calculated from ray theory. Thus, it is more appropriate to think in terms of modes.

Examples of received intensities at the 1-km range are shown in Fig. 3. The received signals shown are the result of pulse compressing the FM transmissions. Two characteristics of these signals are evident: Each individual arrival in a ping has more structure than the transmitted waveform, and there are strong ping-to-ping fluctuations in each arrival. The sidelobes of the transmitted pulse are 180° out of phase with the main lobe, which causes the individual arrivals to have structure of the same time extent as the transmitted pulse, as checked by simulations. The first arrival has most of the total received sound energy. This arrival corresponds to rays that remain in the water column with all initial angles between ±6°. Other arrivals correspond to rays that encounter the bottom, which ray tracing predicts will arrive later by a few milliseconds. Therefore, the fluctuations of total received sound are dominated by the variations in the water column. The depth uncertainty of the sound source was estimated to be within 0.5 m. To estimate the impact of such uncertainty on intensity fluctuations, we correlated the integrated intensity between two receivers vertically separated by 0.9 m. The correlation coefficient is found to be 0.95, thus source vertical motion was ruled out as having substantial impact on intensity fluctuations. There is nothing in our environmental measurements that would deterministically distinguish one pulse from another, so the pulses must be treated as

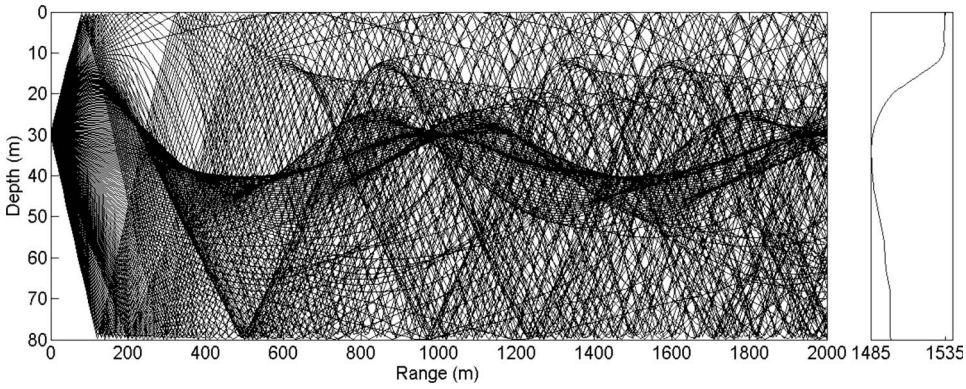


Fig. 2. On the left, rays spanning ±23° with 0.2° increment are given where the source is at 30 m depth. One of the measured sound speed profiles is on the right.

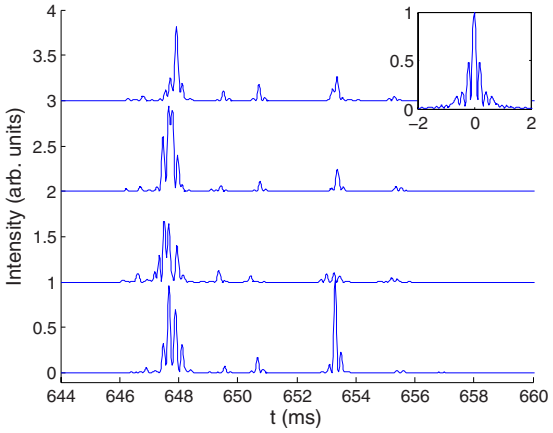


Fig. 3. (Color online) The intensity of sample FM data after pulse compression for receiver at 25 m depth. The time is true arrival time for the bottom trace, others are lined up in time against the bottom trace. The inset is the transmitted signal for reference.

random. The first statistic to consider is the mean intensity. Because the geometrical spreading loss and loss due to seawater absorption are well understood, the term reduced transmission loss will be used for this paper, which is defined as the transmission loss with the cylindrical spreading and seawater absorption removed.

The measured variation in transmission loss (TL) for individual pulses at a fixed range was large and assumed caused by scattering from sound speed variations in the water column. If one considers a sound channel where there is no range-dependent sound speed variation in the water column, so there is no scattering, then the TLs measured at two receivers separated by a small range difference would have similarly large variations as the measurement. The variation in this range-independent case over short distances is due to interference between the modes. However, the measured reduced mean TL is much more constant among all ranges than the individual ping TLs at a fixed range. Thus we infer that the ocean-induced ping-to-ping variability is indicative of random phase addition of multiple modes. This suggests a model in which the mean intensity is the incoherent sum of mode intensities. To complete the model, it is assumed that the mode intensities are, on the average, the same as the range-independent mode intensities. The Dozier and Tappert[5] transport theory makes a similar random phase approximation, but is designed to calculate the mode intensities more accurately.

Sound speed variations measured by the towed chain (such as shown in Fig. 1) were used to quantitatively examine the modal phase changes. The difference between the sound speed fields of two almost coincident tracks from the towed chain is calculated. This difference is integrated, at each depth, over the 1-km segment closest to the acoustic path. The resulting depth variation is then projected onto all 15 modes that remain in the water column at 2 kHz. This simulates the first arrival in Fig. 3. This projection simulates the phase difference for each mode between the two times of the towed chain measurements. These calculations give a standard deviation of 7.4 rads. This value supports the validity of the random phase assumption. The phase difference will increase linearly with increasing frequency and in an undetermined manner with range depending on the correlation of the sound speed fluctuation. This level of fluctuation makes the coherent field small, and supports the assumption that the incoherent mode intensity sum should be used.

The incoherent mode intensity sums at ranges $1-9$ km and for center frequencies $2,3,\ldots,9$ kHz were calculated using KRAKEN.[10] The sound speed profile used in the model was from a CTD cast from the R/V KNORR during the 1-km range measurements and is given in Fig. 2. For the small amount of sound energy interacting with the bottom, the bottom is modeled to be a fluid half space with sound speed 1610 m/s,[9] bulk density 1800 kg/m^3, and attenuation

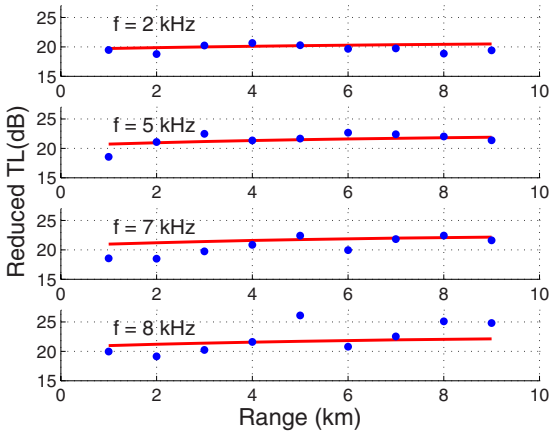


Fig. 4. (Color online) Difference of transmission losses between measurement and model at four frequencies, each with a 100-Hz bandwidth. The data are from the mean of 100 pings and the model is the incoherent sums of all trapped modes.

coefficient 0.2 dB/m/kHz. The last two numbers were decided through iterative fitting. The attenuation in the water column is included in the model. The model transmission losses are compared to those from data in Fig. 4. The reduced TL for all ranges and frequencies is near a constant value of 20 dB. Initially, the wave field spreads spherically, and then transitions to cylindrical spreading. The reduced TL has only cylindrical spreading subtracted, and the spherical spreading leads to extra TL. The range at which the transition from spherical to cylindrical spreading occurs can be estimated from the vertical spacing between nearby rays. This spacing turns out to be 100 m times the initial angular separation of the rays, indicating that the transition occurs at about 100 m, giving an initial 20 dB extra TL.

The model predicts the mean intensity for all ranges and frequencies considered to within a few dB. Because the reduced TL is nearly a constant for all ranges, the acoustic energy is trapped in the sound channel. Error bars were not calculated because successive pings are correlated, and the number of degrees of freedom is yet to be evaluated. Instead, the 100 pings at each station were split into the first and last 50-ping groups and the difference is found to be less than the difference between the model and data. Thus, the data are sufficient to show small inaccuracies of the model assumptions. There are a few places where the model is off by more than 3 dB, for example at 8 kHz and 5 km.

4. Summary and discussion

Mid-frequency sound propagation in shallow water was measured at multiple ranges along with extensive environmental measurements, especially 2D CTD measurements using a towed chain. These measurements provided the opportunity to quantify the statistics of transmission loss. This paper concentrated on estimating the mean intensity, and a companion paper will address the issue of intensity fluctuations. It was found that most of the sound is trapped in the sound channel where water column variability dominated the acoustics field fluctuations. Because of the presence of closely packed caustics at convergence regions, ray theory cannot be used to accurately calculate sound intensity levels. The intensity was estimated from data and compared to an incoherent mode sum model. The comparison indicated that the mean intensity can be modeled to within 2 dB for most of the ranges and frequencies studied, but there were exceptions where the difference between model and data exceeded 4 dB. Although the differences between the model and data are small, they are statistically significant.

Sufficient towed chain data were taken to allow a more detailed statistical model of the internal waves. We expect that Monte Carlo propagation calculations using this statistical model will more accurately describe the data than the simple model presented. The data will

also be used to check the validity of the transport theory,[11] where certain approximations are made. While such a theory, when applicable, can be used to predict the statistics of the intensity, simulations based on the Parabolic Equation method are potentially a more reliable and general way to study mid-frequency acoustics propagation in complex environments.

An important quantity for scientific understanding and applications is the coherent field, where phase coherence over time is measured. However, because the relative positions of the source and receivers were not known to sufficient precision during SW06, field coherence was not studied, but should be a high priority for a future field experiment.

Acknowledgments

This work was supported by the Office of Naval Research. The authors thank the crews of the R/V KNORR and R/V ENDEAVOR for their professional support.

References and links

[1]F. B. Jensen, "Sound propagation in shallow water: A detailed description of the acoustic field close to surface and bottom," J. Acoust. Soc. Am. **70**, 1397–1406 (1981).
[2]M. Siderius, M. Snellen, D. G. Simons, and R. Onken, "An environmental assessment in the Strait of Sicily: Measurement and analysis techniques for determining bottom and oceanographic properties," IEEE J. Ocean. Eng. **25**, 264–386 (2000).
[3]W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey continental shelf," IEEE J. Ocean. Eng. **20**, 321–336 (2000).
[4]T. Duda, J. F. Lynch, A. E. Newhall, L. Wu, and C.-S. Chiu, "Fluctuation of 400 Hz sound intensity in the 2001 ASIAEX South China Sea experiment," IEEE J. Ocean. Eng. **29**, 1264–1279 (2004).
[5]H. DeFerrari, N. Williams, and H. Nguyen, "Focused arrivals in shallow water propagation in the Straits of Florida," ARLO **4**(3), 106–111 (2003).
[6]T. E. Ewart and S. A. Reynolds, "The mid-ocean acoustic transmission experiment, MATE," J. Acoust. Soc. Am. **75**, 785–802 (1984).
[7]T. E. Ewart, "A model of the intensity probability distribution for wave propagation in random media," J. Acoust. Soc. Am. **86**, 1490–1498 (1989).
[8]P. H. Dahl, J. W. Choi, N. J. Williams, and H. C. Graber, "Field measurements and modeling of attenuation from near-surface bubbles for frequencies 1–20 kHz ," J. Acoust. Soc. Am. **124**, EL163–EL169 (2008).
[9]J. Yang, D. Tang, and K. L. Williams, "Direct measurement of sediment sound speed using SAMS in SW06," J. Acoust. Soc. Am. **124**, EL116–EL121 (2008).
[10]M. B. Porter and E. L. Reiss, "A numerical method for bottom interacting ocean acoustic normal modes," J. Acoust. Soc. Am. **77**, 1760–1767 (1985).
[11]L. B. Dozier and F. D. Tappert, "Statistics of normal mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **63**, 353–365 (1978).

Mid-frequency acoustic propagation in shallow water on the New Jersey shelf. II. Intensity fluctuation

Dajun Tang, Frank S. Henyey, Zhongkang Wang,[a] Kevin L. Williams, Daniel Rouseff, Peter H. Dahl, Jorge Quijano,[b] and Jee Woong Choi[c]

Applied Physics Laboratory, College of Ocean and Fishery Sciences, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

djtang@apl.washington.edu, frank@apl.washington.edu, williams@apl.washington.edu, rouseff@apl.washington.edu, dahl@apl.washington.edu

Abstract: The scintillation index and the intensity cumulative distribution function of mid-frequency (2–10 kHz) sound propagation are presented at ranges of 1–9 km in a shallow water channel. The fluctuations are due to water column sound speed variability. It is found that intensity is only correlated over a narrow frequency band (50–200 Hz) and the bandwidth is independent of center frequency and range. Furthermore, the intensity probability distribution peaks at zero for all frequencies, and follows an exponential distribution at small values.

PACS numbers: 43.30.Ft, 43.30.Re, 43.30.Bp, 43.30.Ma [WC]
Date Received: May 8, 2008 **Date Accepted:** July 6, 2008

1. Introduction

During the Shallow Water 2006 Experiment on 11 August '06, a mid-frequency acoustic propagation experiment was performed in the frequency band 2–10 kHz (referred to as the mid-frequency band) and at ranges 1–9 km. The mean intensity has been found to be well modeled by a simple model of incoherent sum of modes.[1] The intensity fluctuation over short time periods is found to be significant and causes uncertainty in predictions of sound propagation. In this work the acoustic arrivals are treated as random and the scintillation index is used for the purpose of quantifying intensity fluctuations.

Mid-frequency sound intensity fluctuations in the deep ocean have been measured[2] and studied by using the method of moments and simulations,[3] where it is found that the intensity fluctuation is small and not saturated. Observations of low-frequency (hundreds of Hertz) intensity fluctuation in the deep ocean have been reported. For example, Duda *et al.*[4] found strong focusing that makes the scintillation index at 400 Hz to be greater than 1. There are some efforts devoted to intensity fluctuations in shallow water at low frequency.[5,6] To the authors' knowledge, research on mid-frequency intensity fluctuation in shallow water is limited.[7–9]

At the experiment site, a typical strong summer thermocline existed, and in addition, warm, salty slope water underlay the colder mid-depth water. Thus, there was a sound speed minimum at approximately 32 m depth, forming a sound channel. The acoustic propagation measurement was carried out along the 80 m, isobath for $1,2,\ldots,9$ km ranges. At each of these stations, 100 identical acoustic transmissions were made in order to study the statistics of sound intensity on single hydrophones. The signals transmitted and studied in this letter are wideband pulses between 20 and 10 kHz. Details of the experiment description can be found in Ref. 1. The letter describes the intensity fluctuation as a function of frequency in terms of scintillation

---

[a]Permanent affiliation: Hangzhou Applied Acoustics Research Institute, 96 Huaxing Road, Hangzhou, China.
[b]Current affiliation: Portland State University, 1900 SW Fourth Avenue, Suite 160–17, Portland, Oregon, 97207.
[c]Current affiliation: Department of Environmental Marine Sciences, Hanyang University, Hanyang, Korea.

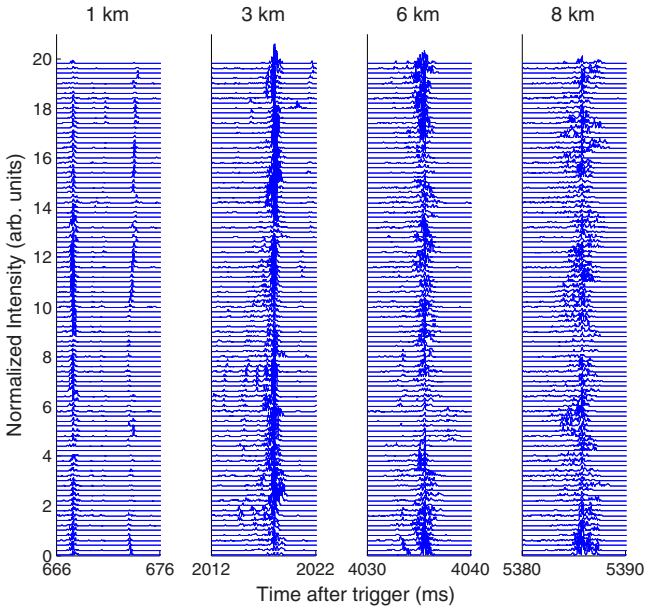


Fig. 1. (Color online) Normalized sound intensity of the 100 individual pings at different ranges. The pings were separated by 20 s. Each successive ping is moved 0.2 from the previous ping on the vertical axis for display. Note that the first arrival at the 1 km range shows deep fades near 9, and 20 along the vertical axis.

index. Cross-frequency correlations of intensity and the cumulative distribution functions of intensity fluctuations are presented.

2. Quantifying intensity fluctuations

On the same day but before the acoustic data at the different range stations were taken, data from a towed conductivity-temperature-depth (CTD) chain over an extended period showed that sound speed close to the sound channel had complex variations,[1] believed to be caused by ambient internal waves and possibly neutrally buoyant thermohaline intrusions. No strongly nonlinear internal waves were observed that day, a special case during the experiment. These ambient internal waves cause intensity fluctuations. These fluctuations are too complicated to be treated deterministically, and necessarily must be dealt with as a random process. The intensity of the 100 pings at selected range stations are shown in Fig. 1, where the ping-to-ping interval is roughly 20 s, and the signal-to-noise ratio is 45 dB after pulse compression of frequency modulated signals. The 10-ms-long windows in Fig. 1 contain at least 97% of the energy of all arrivals for all range stations. Due to space limitations, results from four of the representative range stations are shown; however, results for the range stations not shown have characteristics consistent with those presented in this letter. Because the acoustic source had an unmeasured horizontal motion $O(1\ \mathrm{m})$ relative to the receivers and 0.5 m uncertainty in depth, the data shown in the figure have been pulse compressed and lined up with the first ping at each range. Correlating integrated intensity between two receivers vertically separated by 0.9 m, the correlation coefficient is found to be 0.95, thus source vertical motion was ruled out as having substantial impact on intensity fluctuations.

At the 1 km range, measured sound intensity can be separated into individual arrivals; at other ranges, the arrivals are clustered and not separable. The received pressure can have broadband deep fades, such as the ones shown at the 1 km range in Fig. 1; ping-to-ping fluctuations increase as the range increases. Calculations show that the peak intensity can deviate from its mean value by as much as 12 dB. This letter seeks to quantify the fluctuation.

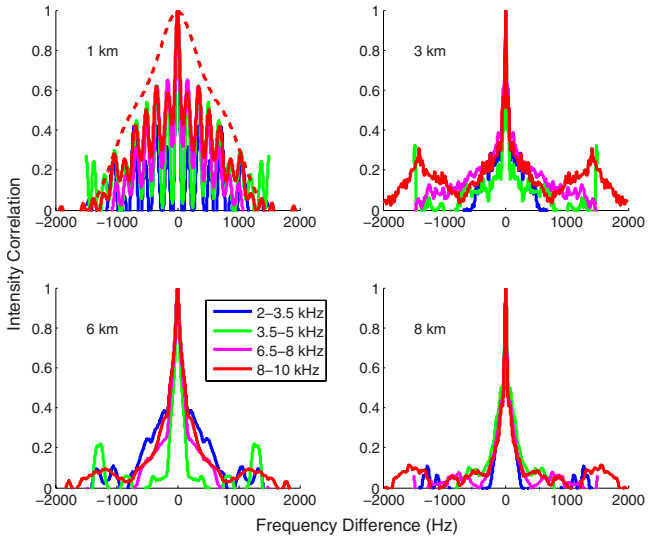


Fig. 2. (Color online) Cross-frequency intensity correlation vs frequency difference in different frequency bands and at different ranges. The dashed curve for the 1 km range is a special case where the first arrivals from each ping in Fig. 1 were used. All other curves used all of the data as given in Fig. 1. The half-power widths for the correlation are 90, 50, 200, and 70 Hz for the 1, 3, 6, and 8 km range.

Intensity fluctuations are often quantified in the scintillation index, SI, defined as SI $=\langle I^2\rangle/\langle I\rangle^2-1$, where I is sound intensity and the angular brackets stand for ensemble average. The intensity, I, can be chosen to have different meanings, thereby leading to different scintillation indices. For example, I could be a time integral, a peak value, or a modal intensity. In both theoretical and simulation approaches dealing with wave propagation in random media, I is naturally chosen to be a single-frequency intensity at a single receiver position; this choice is made here, in order to be able to compare with theoretical and modeled results. A broadband SI is expected to be smaller than the single-frequency SI. If I is associated with a nonrandom process, SI$=0$; if I has a number of statistically independent contributions with random phases (uniformly distributed on $[0,2\pi]$), SI$=1$, and is said to be saturated. In the saturated case, the intensity has an exponential distribution. In order to properly extract the single-frequency SI from data, a suitably small bandwidth needs to be found within which sound intensity is correlated. The appropriate bandwidth can be determined from the cross-frequency intensity correlation. The cross-frequency intensity correlation is a fourth moment of the sound pressure field and was calculated in the following steps. Pressure time series 100 ms in duration were Fourier transformed and squared, resulting in intensity with 10 Hz resolution. For the 100 pings at each range station, the intensity at each frequency bin was normalized by its mean, and then the intensity cross-frequency correlation was calculated. For each lag, δf, the correlation function was averaged over the mean frequency f for which $f+\delta f/2$ and $f0\delta f/2$ are within a chosen band. The bands were chosen to be: 2–3.5, 3.5–5, 6.5–8, and 8–10 kHz.This averaged cross-frequency correlation function is given in Fig. 2 for four ranges.

At the 1 km range, the intensity correlation shows regular oscillations with the same spacing in frequency difference in each frequency band. This oscillation can be qualitatively understood. Note that in Fig. 1, there are two major arrivals separated by 6 ms. The first is the direct arrival that stays in the water column, and the second is a bottom bounce. Assume that the two arrivals have independent scintillation, with scintillation indices S_1 and S_2, both independent of frequency. Also assume that the time spacing T between the arrivals is a constant. The total arrival is written as $a=a_1\exp(-i2\pi fT_1)+a_2\exp(-i2\pi fT_2)$, where $T=T_2-T_1$, then to the lowest nontrivial order the cross-frequency intensity correlation for frequencies f_a and f_b can be found to be

$$\frac{\langle \delta(f_a)\delta(f_b)\rangle}{\langle I(f_a)\rangle\langle I(f_b)\rangle} = S_1 F\left(\frac{\langle a_2\rangle}{\langle a_1\rangle}, f_a\right) F\left(\frac{\langle a_2\rangle}{\langle a_1\rangle}, f_b\right) + S_2 F\left(\frac{\langle a_1\rangle}{\langle a_2\rangle}, f_a\right) F\left(\frac{\langle a_1\rangle}{\langle a_2\rangle}, f_b\right),$$

where $F(R,f)=[1+R\cos(2\pi fT)/1+2R\cos(2\pi fT)+R^2]$, R is the amplitude ratio of the two arrivals, and $\delta(f)=I(f)-\langle I(f)\rangle$. Because the above-presented expression is symmetric with respect to the two frequencies and the dependence of T is periodic, the cross-frequency correlation in the two-dimensional space has a "chessboard" pattern with periodicity of $1/T$. After averaging correlation over the mean frequency within a chosen band, this chessboard pattern becomes a regular oscillation as shown in the 1 km case in Fig. 2, where the periodicity is seen to be $1/T$, as predicted. If the first arrivals were isolated and a correlation was separately calculated, the oscillation disappears and the correlation improves as seen in Fig. 2 given as a dashed curve.

At all ranges longer than 1 km, the arrivals are not separable and are treated as a single arrival. At all frequencies, the intensity cross-correlation function has a sharp peak with a half-power width of 50–200 Hz with no clear range dependency, as shown in Fig. 2. The origin of the narrowness of correlation is yet to be investigated. From the estimated width of the intensity cross-frequency correlation, it is safe to assume that the 10 Hz bin is narrow enough to assure that within the bin the intensity is fully correlated; hence a single-frequency scintillation index can be estimated within each bin. All the intensity values from 2–10 kHz with 10 Hz bin width were divided into the same four frequency bands as when estimating the cross-frequency correlations. In each of the bands, the intensity cumulative distribution function (CDF) $C(I)$ is estimated for each of the nine ranges and is shown in Fig. 3. Here $C(I)$ is defined as the probability that the intensity is less than I. In Fig. 3 we plotted on a logarithmic scale the complementary function, $10\,C(I)$, for easier comparison to the exponential distribution function, which is a straight line on the logarithmic scale. We also plotted numerically determined 95% confidence bounds for the exponential distribution in Fig. 3 by faint dashed curves. If the intensity follows the exponential distribution, data would fall within the bounds. As shown in Fig. 3, the small intensity regime of $C(I)$ for all frequencies and ranges follows the exponential distribution. The probability distribution function (PDF) is the derivative of $C(I)$, and, for an exponential distribution, is largest at zero, thus the intensity PDF peaks at zero. This intensity fluctuation can also be observed in the ping-to-ping variations in intensity as in Fig. 1, where the first arrivals at the 1 km range show deep fades at several places. As there were no strongly nonlinear internal waves on the day of the measurements, one would expect the intensity PDF to peak around the mean intensity and diminish at small intensities. The observation that the intensity at small values follows the exponential distribution for a wide band of frequencies is to us a surprising result. Saturation happens when a large number of independent arrivals contribute additively to the signals. However, our observations show that even at the 1 km range, the small intensities follow the exponential distribution, even though at such short range it is not expected that the arrivals are saturated. The large intensity values are understood to be caused by focusing effects. Clearly they do not follow the exponential distribution but they constitute low probability. Therefore, we conclude that the intensity fluctuation is not saturated, and the reason needs to be investigated.

3. Summary and discussion

Mid-frequency (2–10 kHz) sound intensity fluctuations at 1–9 km range stations were investigated through the single-frequency scintillation index. The intensity fluctuations are due to water column sound speed variability from ambient internal waves; the data were collected on a day when there were no strongly nonlinear internal waves present. It is highly unlikely that fish scattering is the cause of the fluctuation because the observed large fluctuation requires high fish density. Also, such fish presence would result in noticeable attenuation in the mean intensity versus range, which is not observed.[1]

Two findings were not anticipated before the experiment: intensity is only correlated over a narrow frequency band (50–200 Hz) and the bandwidth is independent of center frequency and range; the intensity PDF peaks at zero for all frequencies, and follows an exponen-

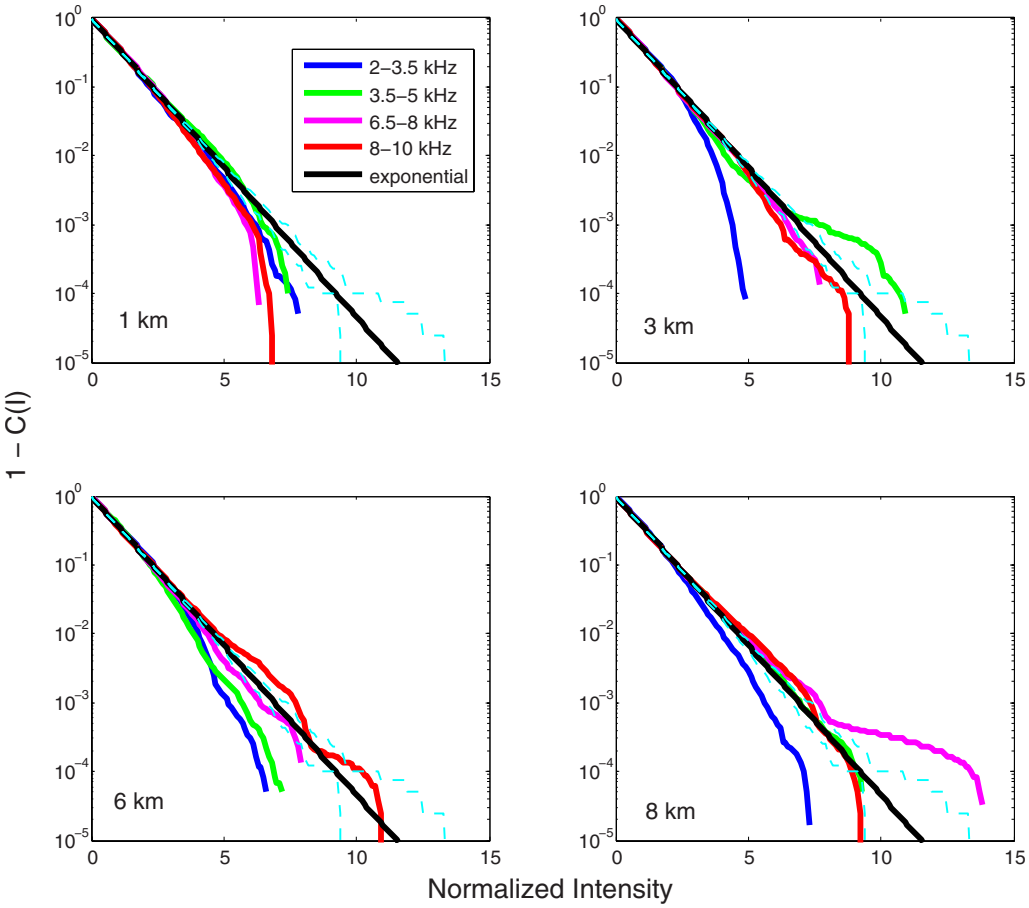


Fig. 3. (Color online) $1-C(I)$, where $C(I)$ is the cumulative distribution function (CDF), at different range vs normalized intensity at different range and frequency bands. The exponential distribution is also plotted for reference. The faint dashed curves are the 95% confidence bounds for the exponential distribution given the number of samples in the experiment.

tial distribution at small values. What causes the observed deep fades in intensity? Such deep fades manifest as uncertainty in predicting mid-frequency sound propagation in shallow water. To understand the intensity fluctuation in order to estimate uncertainty, two potentially profitable approaches are numerical simulations based on the parabolic equation approximation and transport theory such as the one proposed by Dozier and Tappert,[10] where the validity of some key assumptions needs to be investigated.

Acknowledgment

This work was supported by the Office of Naval Research.

References and links

[1]D. Tang, F. S. Henyey, Z. Wang, K. L. Williams, D. Rouseff, P. H. Dahl, J. Quijano, and J. W. Choi, "Mid-frequency acoustic propagation in shallow water on the New Jersey shelf. I Mean intensity," J. Acoust. Soc. Am., **124**, EL85–EL90 (2008)
[2]T. E. Ewart and S. A. Reynolds, "The Mid-Ocean Acoustic Transmission Experiment, MATE," J. Acoust. Soc. Am. **75**, 7850–7802 (1984).
[3]T. E. Ewart, "A model of the intensity probability distribution for wave propagation in random media," J. Acoust. Soc. Am. **86**, 1490–1498 (1989).
[4]T. F. Duda, J. F. Lynch, A. E. Newhall, L. Wu, and C. Chiu, "Fluctuation of 4000 Hz sound intensity in the

2001 ASIAEX South China Sea Experiment," IEEE J. Ocean. Eng. **99**, 1264–1279 (2004).
[5]A. Fredricks, J. A. Colosi, J. F. Lynch, G. Gawarkiewicz, C.-S. Chiu, and P. Abbot, "Analysis of multipath scintillations from long range acoustic transmissions on the New England continental slope and shelf," J. Acoust. Soc. Am. **997**, 1038–1057 (2005).
[6]B. Pasewark, S. Wolf, M. Orr, and J. Lynch, "Acoustic intensity variability in a shallow water environment," Acoustic Variability 2002, SACLANT Undersea Research Center, pp. 11–18.
[7]J. A. Simmen, S. M. Flatté, H. A. DeFerrari, H. Nguyen, and N. J. Williams, "Near-caustic behavior in a 2700 km acoustical experiment," J. Acoust. Soc. Am. **105**, 3231–3244 (1999).
[8]H. A. DeFerrari, N. J. Williams, and H. Nguyen, "Focused arrivals in shallow water propagation in the Straits of Florida," ARLO **4**, 106–111 (2003).
[9]J. E. Lyons, D. L. Bradley, R. L. Culver, and K. M. Becker, "The effect of medium variability on acoustic variability: Internal waves and thermohaline intrusions (spice)," J. Acoust. Soc. Am. **119**, 3428 (2006).
[10]L. B. Dozier and F. D. Tappert, "Statistics of normal mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **69**, 353–365 (1978).

Observed limiting cases of horizontal field coherence and array performance in a time-varying internal wavefield

Jon M. Collis
Boston University, 110 Cummington St., Boston, Massachusetts 02215
jcollis@bu.edu

Timothy F. Duda and James F. Lynch
Applied Ocean Physics and Engineering Department, MS 11, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
tduda@whoi.edu, jlynch@whoi@edu

Harry A. DeFerrari
University of Miami, Miami, Florida 33149
hdeferrari@rsmas.miami.edu

Abstract: Using a moored source and horizontal/vertical line array combination, horizontal coherence properties of high signal to noise ratio ($\geq$20 dB) 100–1600 Hz signals have been measured. Internal waves in the area of the measurement created moving episodic sound-speed anomaly structures, influencing coherence length. Measured horizontal coherence scales for 100 Hz ranged from 5 to 20 acoustic wavelengths, and were inversely related to the sound-speed anomaly strength. Horizontal field properties were compared with fields computed using modal decompositions of the vertical signals. The comparison allows azimuthal field coherence properties to be studied apart from normal-mode interference effects.

PACS numbers: 43.30.Re, 43.30.Bp, 43.30.Ft [DKW]
Date Received: February 26, 2008 **Date Accepted:** May 14, 2008

1. Introduction

To study acoustic fields and their coherence over significant distances, the research community has recently taken steps to enable routine use of co-located horizontal and vertical line arrays (HLAs and VLAs), typically arranged in the shape of the Roman letter L, with the horizontal section on the seafloor.[1] To analyze horizontal coherence we introduce a procedure tailored to oblique incidence on the HLA. Mode interference that occurs over range will introduce incident-angle dependent signal variability along the HLA which we seek to separate from transverse signal decorrelation effects, associated with azimuthally varying propagation (i.e., without cylindrical symmetry). The procedure uses joint analysis of the vertical and horizontal line array receptions. Impulsive mode arrivals (including mode multipath) from fixed sources are estimated using the vertical array. Coherence of measured waveforms of oblique incidence on the horizontal array is compared with that of signals having spatially independent normal mode amplitudes and wave numbers (the fixed-mode field) synthesized using the vertical array modal arrivals. The comparison discloses azimuthal (relative to the source) decorrelation effects without the restriction of having broadside HLA signal incidence. Array gain degradation results for the measured and synthetic fields are also presented, providing additional estimates of field coherence. The results show two behaviors, one where the azimuthal effect is significant enough to be measureable, and one where it is not. These correspond roughly (but not exactly) to conditions of relatively short and long horizontal coherence scales, and to conditions of large and small array gain degradation.

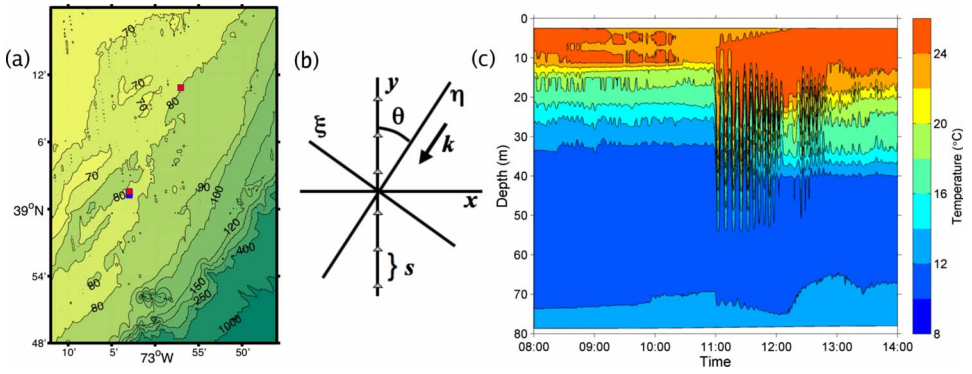


Fig. 1. (Color online) (a) Depiction of the receiver array (southern red dot) and source (northern red dot) moorings. The horizontal leg of the array is aligned very close to north/south. The bearing to the sound source is approximately 26.2° at the southern (VLA) end of the array. (b) Plane view defining acoustic angle of incidence and axis orientations. (c) Contour plot of isotherms measured by a mooring near the acoustic source. Contour intervals are 1.5 C.

2. Data and methods

Data from a 48-element L-array deployed during the Shallow-Water 2006 (SW06) field program[2] are used here to measure coherence. As part of the experiment, numerous acoustic and oceanographic moorings were deployed, with the majority following a "T" shaped geometry. The upright leg of the T had an along-shelf alignment following (approximately) the 80 m isobath, at heading 30° (true), whereas the other was across shelf from depths of 58–500 m (true heading 300°). The L-array was located near the intersection of the T [Fig. 1(a)] and had two sections: A 16-element VLA spanning four-fifths of the water column (water depth 79 m) and a 32-element HLA, 472 m in length, aligned south to north. Moored acoustic sources were placed near the far end of the along-shelf line and near the inshore end of the cross-shelf line, to provide known signals.

The signals considered here are 100 Hz pulses produced by the "Miami Sound Machine" source, located 19.2 km from the L-array [Fig. 1(a)]. The heading from the HLA center to the source was 26.2° [incidence 63.8° from broadside, Fig. 1(b)]. Each half hour the source emitted a continuous series of broadband phase-coded pulses at a center frequency of 100 Hz (90 s, 36 pulses), then switched to higher frequencies. The received signals were replica correlated[3] to create the data shown here. Internal ocean state was measured with moored current meters and thermometers at locations near the source, the receiver, and midway between.

2.1 Synthesized and measured horizontal fields

The emphasis of our technique is to evaluate horizontal coherence in planes transverse to the acoustic propagation direction. This variability can be measured directly with an HLA only for signals at broadside incidence, a limiting geometry for experimental purposes. Here, VLA data are incorporated into the analysis to account for expected range variability of obliquely incident HLA signals.

For this analysis, a synthetic acoustic field is computed from the modal decomposition of each pulse measured along the VLA, and used as a metric for comparison with the data. This procedure separates two spatial decoherence effects. The first effect is caused by mode interference, which degrades coherence if the modes are not fully separated via dispersion and incidence is not broadside. This limits coherence of signals sampled along the array in a manner that is dependent on incident angle. The second effect is caused by azimuthally dependent propagation from the source, which would be evident when the projected array aperture normal to the source direction is large compared to the coherence length in the transverse plane (normal to propagation direction).

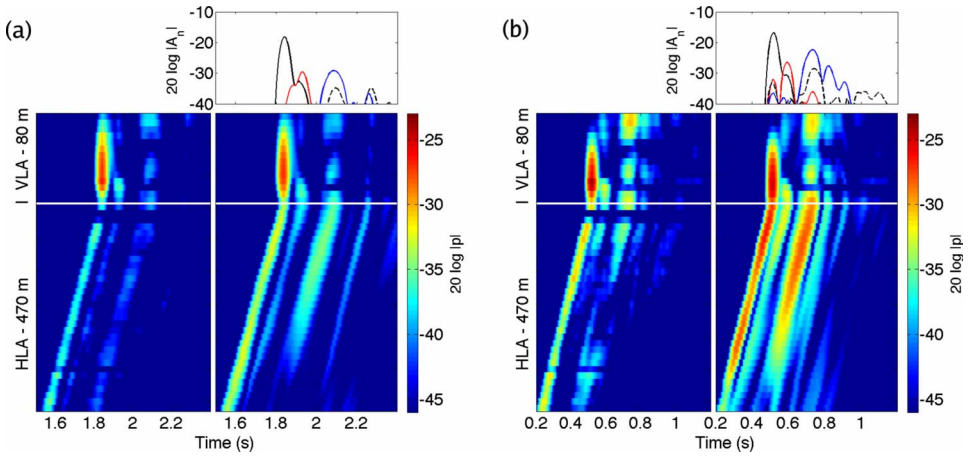


Fig. 2. (Color online) Intensity (measured and synthesized) and mode coefficients along the L-array during different regimes of internal wave activity. The result of the modal decomposition is shown above the synthesized field. (a) Weak scattering period. (b) Strong scattering period.

First, the data are match-filter processed to obtain the complex acoustic pressure, $p_d(\mathbf{r_j},t)$, $\mathbf{r_j}=(x_j,y_j,z_j)$, at each element ($j$) along the L array (the subscript d denotes measured data). Sound speed profiles are determined using temperature data collected along the VLA, and a nearby buoy for the near surface. From these, time-dependent normal mode functions and associated wave number values are determined using a simplified geoacoustic model (discussed below) of the site. Mode filtering is applied to p_d along the VLA to obtain time series of mode content. The acoustic signal at the VLA element positions can be written in terms of normal modes, $p_d(x_0,y_0,z,t)=\Sigma_{n=1}^{N}A_n(t)\phi_n(z,t)$ where p_d is the complex matched-filter output, A_n are the mode coefficients, ϕ_n are the vertical mode functions, and x_0, y_0 locate the VLA at the origin. There are 16 VLA hydrophones and at 100 Hz only 4–6 modes are expected to arrive. Because this system is well determined, the pseudoinverse method[4] is used to obtain the A_n. A 9-layer bottom model is used in the mode computations, in which geoacoustic and stratigraphic properties have been based on past studies conducted at the experimental site.

Assuming the ocean does not change over the horizontal extent of the HLA, the mode wave numbers, amplitudes, and relative mode-shape amplitudes at the seafloor can then be used to synthesize the fixed-mode field along the bottom-resting HLA (or anywhere else). The synthesized field along the HLA is given by

$$p_S(x_0,y-y_0,z_b,t)=\sum_{n=1}^{N}A_n(t)\phi_n(z_b,t)e^{ik_n\Delta\eta}, \tag{1}$$

where the η direction in an (η,ξ) coordinate system is toward the source [Fig. 1(b)] and the subscript S denotes synthetic data. Because the HLA is defined to extend in the y-direction (north in this case), $\Delta\eta=\cos\theta(y-y_0)$, where incident angle θ is defined in Fig. 1(b), and where $z=z_b$ is the seafloor depth.

Figure 2 shows measured and synthesized intensity time series during times of weak and strong scattering effects. The times are 08:30 and 13:30 and the sound speed condition measured at those times can be deduced from Fig. 1(c). In each of Figs. 2(a) and 2(b), the measured intensity is shown on the left, four extracted mode coefficient series $A_n(t)$ are shown above, and to the right is $P_S(t)$ computed using a four-term modal sum. The synthetic fields show expected nonbroadside interference patterns for the fixed-mode situation. For the weakly scattered case [Fig. 2(a)], modes one to four are largely temporally dispersed and resolvable, and the synthetic and measured fields agree for the low modes but differ from each other for the

higher-order, 3rd and 4th modes. For the strong scattering case [Fig. 2(b)], the mode arrivals are multiple and extended (spread) in time, and the fields along the HLA differ strongly from each other.

2.2 Coherence lengths

The field p_d will possess a characteristic scale L_d along the array, defined as the point where coherence (normalized spatially lagged cross-correlation)

$$R_d(\delta) = \mathrm{Re}\left\{ \left\langle \left\langle \frac{P_d(y) \circ P_d^*(y+\delta)}{\left(\sum_t |P_d(y)|^2 \sum_t |P_d(y+\delta)|^2 \right)^{1/2}} \right\rangle_A \right\rangle_E \right\} \tag{2}$$

is reduced to some specified fraction of its zero-lag ($\delta=0$) value, where the brackets $\langle\cdot\rangle_E$ indicate ensemble averaging (in this case averaging over many pulse arrivals), the brackets $\langle\cdot\rangle_A$ indicate averaging along the length of the array (varying y), the $\circ$ denotes the Hadamard (element by element) product, and $P_d(y)$ is the time series $p_d(y,t)$. The time series is of 0.5 s duration. Only the zero time-lag value of a two-dimensional correlation is used, with p_d adjusted to arrive at time zero at each element (beam steering). The nonlinear shape (distortion) of the array, obtained via analysis of fully time dispersed normal modes, is accounted for, with the measured distortion projected onto the $\theta \approx 26.2°$ beam angle causing phase errors of up to 2 radians. The coherence scale L_S of the synthetic field is similarly obtained via $R_S(\delta)$, which is defined analogously to R_d except that p_S is used in place of p_d. Comparing L_d to L_S allows detection of signal fluctuations with smaller coherence scales than resulting from modal interference in the synthetic (fixed A_n and k_n) field. This would indicate azimuthally dependent propagation. Comparing the synthetic and data fields for these situations may divulge the responsible scattering effects.

Array gain calculations can also be used to estimate effective coherence lengths.[5] Under the assumption that noise is uncorrelated between the sensors, the gain degradation can be related to signal coherence length. Signal gain G_c is calculated as

$$G_c(M) = \frac{S^2}{N^2}, \quad S = \left\langle \overline{\frac{1}{M}\sum_{j=M} p_j(t)} \right\rangle_{\mathrm{sa}}, \quad N = \left\langle \overline{\frac{1}{M}\sum_{j=M} p_j(t+\Delta t)} \right\rangle_{\mathrm{sa}}, \tag{3}$$

where $\langle\rangle_{\mathrm{sa}}$ denotes an ensemble average over independent sub-apertures, and the overbar indicates an arithmetic mean in time over the duration of the pulse, S is the received signal, N is the background noise, Δt is a time chosen such that the p_j do not include signal, and $M=2^n$ where $n=0$ to 5 is the number of elements in six HLA subaperture sizes. The time series t for both S and N calculations is of 0.4 s duration.

Theoretical $G_c(M)$ curves can be computed for fields with specific coherence properties and uncorrelated noise. Using the notation of Refs. 5 and 6, let the correlation function be $R(ns)=\rho^n=\exp(-(\alpha ns/\lambda)^p)$, with s being the inter-sensor distance, n a positive integer, λ the acoustic wavelength, α a parameter, and the exponent $p=1$. Computed values of $G_c(M,\rho) = (1+\rho)/(1-\rho) - (2\rho(1-\rho^M))/(M(1-\rho)^2)$ for various ρ (and thus various α) under these assumptions are presented in Figs. 3(a) and 3(b) (smooth curves).[6] Note that a characteristic coherence length is given by $L_h=\lambda/\alpha$, which is easily obtained from ρ. Also note that using a different exponent p will result in analogously different characteristic coherence lengths.

3. Results

Applying the new technique, coherence properties of transmissions from three 6 hour time periods are computed and shown here. The time periods are depicted in Fig. 3(c) which shows symbols at the arrival times of analyzed pulses, along with the signal to noise ratio of the pulses. The pulses are divided into two groups, those with strong p_d, p_S and $R_d(\delta), R_S(\delta)$ similarities (case 1), showing behavior as in Fig. 2(a), and those without, as in Fig. 2(b) (case 2). Also shown

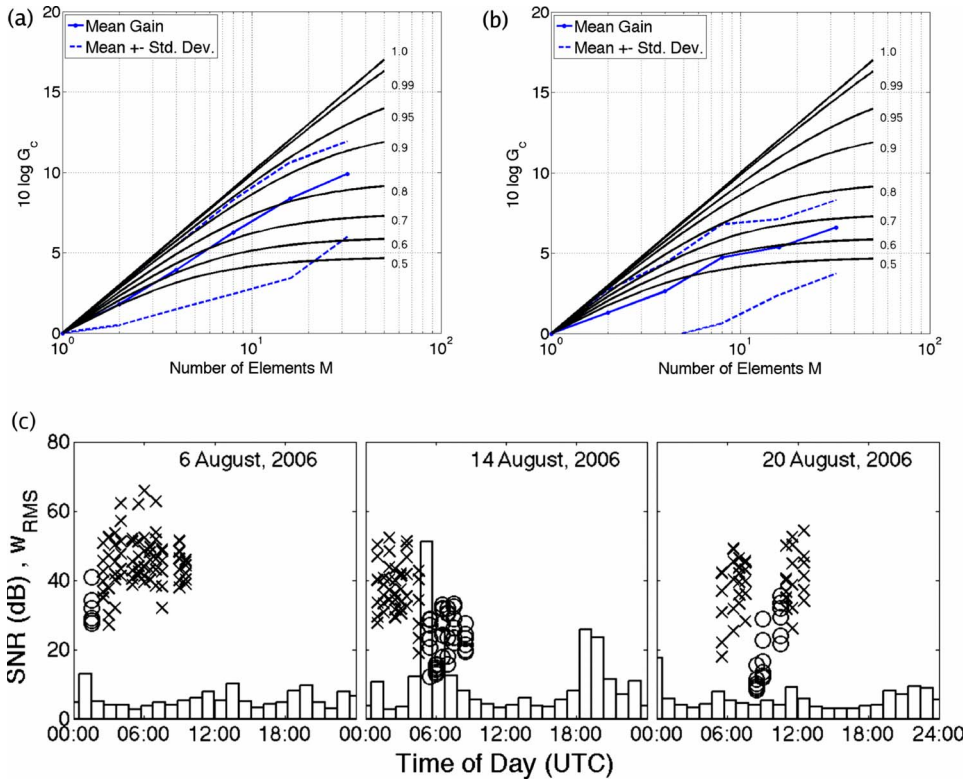


Fig. 3. (Color online) (a) and (b) Array signal gain for the measured data calculated using an ensemble average as a function of aperture size, . (a) Weak scattering time periods. (b) Strong scattering time periods. (c) Plots showing signal to noise ratios for the acoustic transmissions considered in the limiting case coherence plots. Short coherence lengths are indicated by circles and long coherence lengths by x's. Superimposed on the the plot are scaled (20 times), hourly averaged values for the RMS vertical velocity (cm/s).

in Fig. 3(c) are RMS vertical velocity values (scaled by a factor of 20) at one mooring near the HLA, averaged over 1 hour time windows, indicating internal ocean activity, which exhibits some correlation with both the signal to noise ratio and the scattering categorization. Figure 4 shows the correlation functions R_S and R_d for each of the cases, obtained by averaging the correlation functions for individual pulse arrivals.

For the case 1 data, the R curves are consistent with interference patterns of fixed-mode fields. For the case 2 data, the coherence scale from the R_d curve is reduced below that of the R_S via azimuthally dependent propagation and scattering effects. From the curves, correlation scales are defined by where they reach e^{-1} (noting that they can remain high by effects such as coherence in the noise field, for example). For case 1, L_S=220 and L_d=250 m, giving good agreement. For case 2, L_S=140 and L_d=80 m.

Returning to Figs. 3(a) and 3(b), these show $G_c(M)$ computed from the data. The curves show mean results for pulses (plotted with plus/minus one standard deviation) in each of the two groups. The curves can be used to infer ρ and L_h values. These correspond to ρ=0.9, L_h=150 m (case 1 period), and to ρ=0.8, L_h=60 m. Note that there is a mismatch in the array gain degradation curves between the actual theoretical shape and the shape obtained from the data. Possible explanations for this mismatch are that the function $\exp(-ns/H\lambda)$ may be too simple for this data or that an aspect of the data is nonstationary over the ensemble. Because the shapes of the curves do not exactly match, the asymptotic value of each array gain curve is used to obtain ρ (as opposed to selecting ρ from a theoretical curve that is close to the data curve).

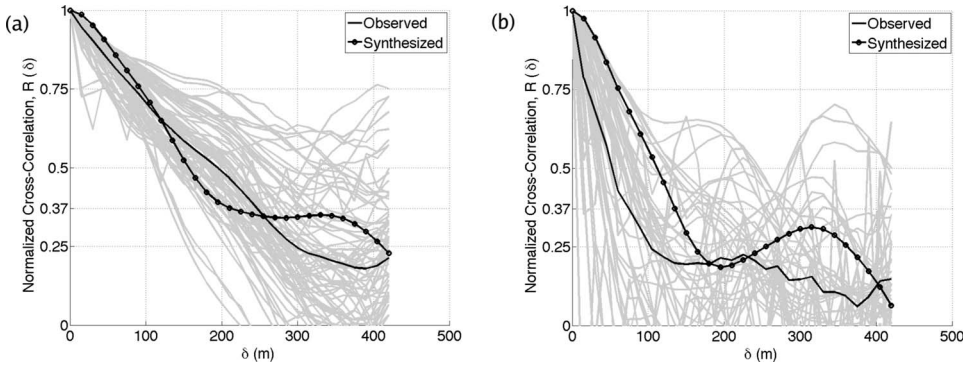


Fig. 4. Relative acoustic coherence along the horizontal line array during the limiting case times. The measured acoustic field (solid curve) is compared against the synthesized field (solid dotted curve). The curves are taken to be the average of individual transmissions (displayed in gray in the plot). (a) Long coherence length (weak scattering) time periods. (b) Short coherence length (strong scattering) time periods.

4. Summary

A method to investigate horizontal coherence of acoustic fields consisting of many normal modes arriving at an L array with nonbroadside geometry has been presented here and applied to a measured data set. The method allows interference of a homogeneous mode structure to be distinguished from horizontal de-correlation from azimuthally varying adiabatic or coupled mode propagation. Coherence length (L_h) results agree with results from array-gain degradation analysis. In addition to providing L_h estimates for given signal incidence angle, the method detects signatures of scattering behavior that can limit array effectiveness at broadside incidence.

Broadband 100 Hz signals along a fixed path during time periods having weak internal variability were observed to have clearly separated modal arrivals and negligible mode multipath effects. In contrast, signals during a period having strong internal variability divulged strong scattering effects and spread-out, multiple mode arrivals. During the period of weak internal ocean activity, L_h was of order 15λ, with estimates from the HLA data and from a synthesized fixed-mode field agreeing. This implies that mode interference sets the horizontal correlation scale, and that the scale will be dependent on incident angle (equivalently, to the source-receiver range difference over the HLA). For our particular geometry, interference controls L_h, so that adjustment to a more-broadside alignment will increase the coherence scale along the array until a limit (not known from this processing) is reached. During the period of strong internal ocean activity, L_h was observed to be shorter, 5λ for the data and 10λ for the synthetic result, consistent with expectations. The shorter result for the data implies that azimuthally varying propagation can degrade array performance. The useful (coherent) aperture for even broadside geometry would be limited, in this case to the projection of the data result onto a plane transverse to the propagation path (i.e., $5\lambda \cos(26.2^\circ) \approx 4.5\lambda$).

Thus, two shallow-water propagation regimes are observed: One where near-broadside HLA geometry may have very long horizontal coherence length and high array gain, and one where it cannot. These analyses will be extended to the entire month-long time series of signals at other frequencies, including those traveling cross shelf. The effect on the results of finite noise level and noise correlation along the array must also be evaluated.

Acknowledgments

The authors acknowledge the use of array processing codes written by Arthur Newhall and Ying-Tsong Lin. We thank Hien B. Nguyen for helping with analysis of signals from the Miami

source. This work was funded by grants to Boston University and the Woods Hole Oceanographic Institution from the Ocean Acoustics Program at the U.S. Office of Naval Research, including an ONR Postdoctoral Fellowship award to the first author.

References and links

[1]M. H. Orr, B. H. Pasewark, S. N. Wolf, J. F. Lynch, T. Schroeder, and C.-S. Chiu, "South China Sea internal tide/internal waves - Impact on the temporal variability of horizontal array gain at 276 Hz," IEEE J. Ocean. Eng. **29**, 1292–1307 (2004).

[2]A. E. Newhall, T. F. Duda, K. von der Heydt, J. D. Irish, J. N. Kemp, S. A. Lerner, S. P. Liberatore, Y.-T. Lin, J. F. Lynch, A. R. Maffei, A. K. Morozov, and A. Shmelev, "Acoustic and oceanographic observations and configuration information for the WHOI moorings from the SW06 experiment," Woods Hole Oceanographic Institution, WHOI Tech. Report No. WHOI-2007-04, Woods Hole, MA, 2007.

[3]H. DeFerrari, N. Williams, and H. Nguyen, "Focused arrivals in shallow water propagation in the Straits of Florida," ARLO **4**, 106–111 (2003).

[4]C. T. Tindle, K. M. Guthrie, G. E. J. Bold, M. D. Johns, D. Jones, K. O. Dixon, and T. G. Birdsall, "Measurements of the frequency dependence of normal modes," J. Acoust. Soc. Am. **64**, 1178–1185 (1978).

[5]W. M. Carey, "The determiniation of signal coherence length based on signal coherence and gain measurements in deep and shallow water," J. Acoust. Soc. Am. **104**, 831–837 (1998).

[6]H. Cox, "Line array performance when the signal coherence is spatially dependent," J. Acoust. Soc. Am. **54**, 1743–1746 (1973).

Temporal coherence of mode arrivals

Harry A. DeFerrari
Division of Applied Marine Physics, RSMAS—University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149
hdeferrari@rsmas.miami.edu

James F. Lynch and Arthur Newhall
Applied Ocean Physics and Engineering Department, MS 11, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Abstract: Temporal coherencies are compared for individual pulse arrivals of 100 and 800 Hz centered broadband signals propagating through identical shallow channels during periods of low and high internal wave energy. All 100 Hz modes are clean, distinct in time, and remarkably coherent ($t_{coh} > 1$ h). A near continuum of modes is observed for the 800 Hz signal with much shorter coherence times. Higher order modes are less coherent. Internal wave scattering appears to dominate the 100 Hz signals, whereas bottom scattering also randomizes the 800 Hz signal. Observations are consistent with similar measurements for this and other sites for intermediate frequencies.

PACS numbers: 43.30.Re [WC]
Date Received: April 17, 2008 **Date Accepted:** June 9, 2008

1. Introduction

Many estimates of temporal and spatial coherence use a lagged covariance calculation of the entire received signal. In such methods, multipath interference variability is often folded into coherence estimates. Here, temporal coherencies are estimated for temporally separable arrivals that are associated with individual acoustic modes of propagation, thus eliminating multipath interference contributions.

Temporal coherence has been studied during three previous experiments at the Shallow Water Experiment 2006 (SW06) site and several other sites around the world. An excellent summary paper (Yang, 2006) compiles the results of these studies for low and intermediate frequencies (200 Hz to 3 KHz) and for ranges out to 40 km. The studies are extensive, compiling data for a great variety of environmental conditions. But most of these estimates of temporal and spatial coherence use a lagged covariance calculation of the entire received signal and do not look at individual modal (or ray) pulse arrivals. The latter is a sensible approach if one is concerned with understanding cw variability in shallow water.

In DeFerrari (2008), individual mode coherencies were computed for a site in the Florida Straits. Most relevant to this study are the findings for midfrequencies (200–800 Hz). Specifically, decorrelation times decrease with frequency and with increasing mode number at a given frequency—by as much as a factor of 4 for the latter. Further, single arrivals/modes are as much as ten times more coherent in time than the combined multipath signal.

The data presented here are selected to examine the extreme conditions during periods of quiescent internal wave activity with low frequency signals that give stable modes with long coherence times, as compared to the upper limit of active internal waves with higher frequencies and, thus, unstable modes that decorrelate quickly. The combined results of these studies give a consistent picture of the frequency and modal dependence of temporal coherence even though the data are for different sites.

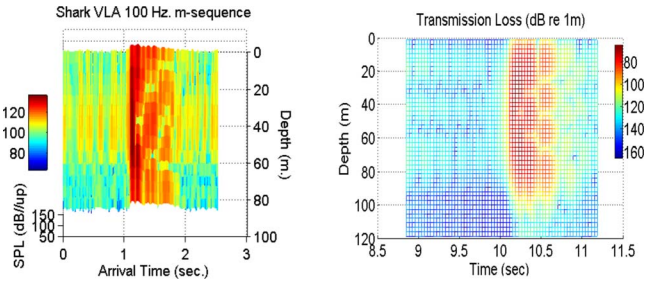


Fig. 1. (Color online) PE model prediction of pulse response with depth compared with observations with the VLA of SHARK at a range of 19.2.

2. The measurements

Our focus here is on the *m*-sequence receptions at three of the single hydrophone receiver units (SHRUs), and the vertical line array receptions with the SHARK receiver. The acoustic source was the Miami Sound Machine (MSM), which transmits *m* sequences at each of five center frequencies, 100, 200, 400, 800, and 1600 Hz, each with a 25% bandwidth (Nguyen *et al.*, 1996). The source was moored in 79 m of water and the transmissions were continuous for a 50 h period, switching between frequencies at 2 h intervals. The signals were received with single hydrophone receiver units, SHRUs 51, 52, and 53, located at ranges of 21.28, 24.57 and 17.12 km, at depths of 85, 107, and 82 m, respectively. The resulting data have negligible source motion and are well suited for temporal coherence calculations. The SW06 acoustic propagation measurements have been described previously in more detail (Collis *et al.*, 2008).

3. Modes and arrivals

The mode/travel time relationship is illustrated in Fig. 1, where the SHARK vertical line array receptions are compared with potential energy (PE) predictions for the 100 Hz *m*-sequence signals after pulse compression. The correspondence between the modes and the arrivals in time is clear. The cutoff of the mode arrivals after mode 4 is abrupt owing to strong bottom loss for modes above the critical angle.

4. Oceanographic fluctations

The site of the MSM-SHRU's transmissions is near, and roughly parallel to, the continental shelf break. This shallow area is inside of a retrograde front that blocks the transmission of offshore internal waves up onto the shelf. The principle cause of fluctuations of the sound speed field is locally generated nonlinear internal waves that propagate shoreward. The internal waves are generated by spilling of tidal currents over the shelf edge. As a result, periods of intense internal wave activity are episodic and timed with local tides. Figure 2 shows the temperature–depth record from an 11 thermistor vertical array located near the MSM. The record is 50 h long, corresponding to the time of continuous transmission and SHRU reception. The pulse responses for two 2 h periods are shown for the 100 Hz data in Fig. 2 (top panel). A strong thermocline exists early and late in the records and nearly isothermal conditions persist for a few hours during the middle period. Of course, internal waves have much greater effect on sound speed when a thermocline is present.

The potential energy of the internal wave field can be estimated from temperature records by the follow relationship (Gill, 1982):

$$\mathrm{PE} = (\rho/2)\,\eta^2 N^2, \tag{1}$$

where N is the buoyancy frequency and

$$\eta = T'/dT/dz. \tag{2}$$

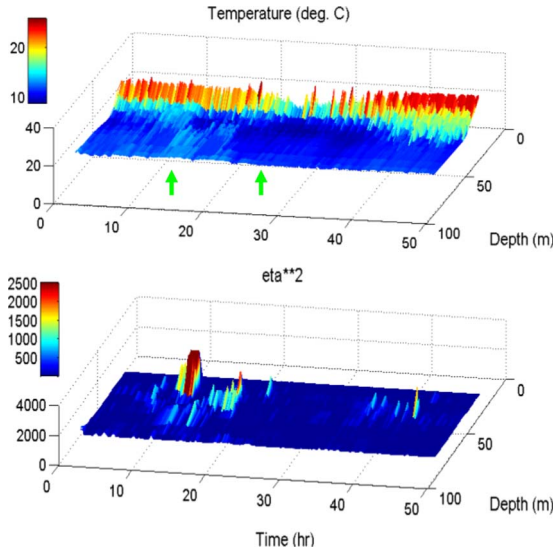


Fig. 2. (Color online) (Top) Temperature–depth time series during 50 continuous transmissions from MSM to SHRU receivers, (Bottom) η^2, proportional to internal wave energy. Green arrows indicate time of 2 h sample used for Fig. 3.

η depends on the temperature perturbations and gradient and is readily computed with the temperature–depth recordings. η squared, proportional to the potential energy of internal wave field, is plotted below the temperature data of Fig. 2. Clearly, there are bursts of high energy internal waves during the first and last third of the record with a quite low energy period in between. This fortuitous sequence of events allows observation of acoustic temporal coherence during periods of high and low internal wave energy.

5. Temporal coherence

Temporal coherence is a statistical measure of the change of a wave form with time. It is a complex quantity that depends on both the amplitude and phase. Details of the calculation method have been described previously (DeFerrari 2008) and are briefly repeated here for convenience. The coherency formula of Eq. (3) is a time lagged normalized covariance calculation,

$$\mathrm{COH}(t,\tau) = \frac{\langle (p(t)p(t+\tau))^2 \rangle_{\Delta t,\Delta T}}{\langle p(t)^2 \rangle_{\Delta t,\Delta T} \langle p(t+\tau)^2 \rangle_{\Delta t,\Delta T}}, \tag{3}$$

where $p(t)$ is the channel pulse response, computed by pulse compression for a 90 s average (τ) of the received m sequences. Approximately 2 h (ΔT) of pulse responses are stored for the coherency calculation. The normalized lagged cross product is repeated for every section (Δt) of $p(t)$. The result is a coherency for every $\Delta t = 4/f$ wide section of the received signal as shown in Fig. 3. Each estimate of COH has 48 degrees of freedom.

Pulse response time histories are plotted, Fig. 3, for a 2 h segment of the 100 Hz receptions during a time of intense internal wave activity, indicated by the left green arrow of Fig. 2. A second sample, 10 h later, during a relatively quiet time is indicated by the green arrow to the right. The temporal coherence computations for each sample are plotted directly below the pulse responses. Similar results are presented for the 800 Hz transmission in Fig. 4.

6. Discussion of results

If one assumes a specular reflection from the ocean surface, there are only two possible causes of mode distortion and scattering; interactions with the seabed (McDaniel, 1982; Evans, 1983), and interactions with the internal wave field (Duda and Preisig, 1999; Preisig and Duda, 1997).

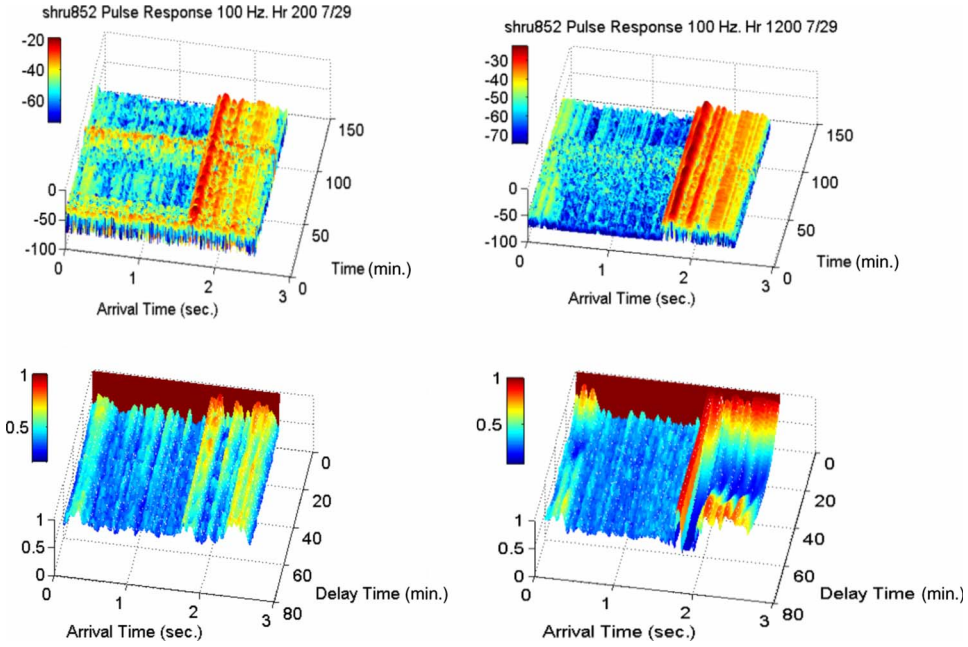


Fig. 3. (Color online) (Left top) Time history of the 100 Hz signal pulse responses at the time sample (arrow—left) of Fig. 2. (Left bottom) Coherency. (Right) Time history of 100 Hz signal pulse responses for sample time (arrow—right) of Fig. 2. (Right bottom) Coherency.

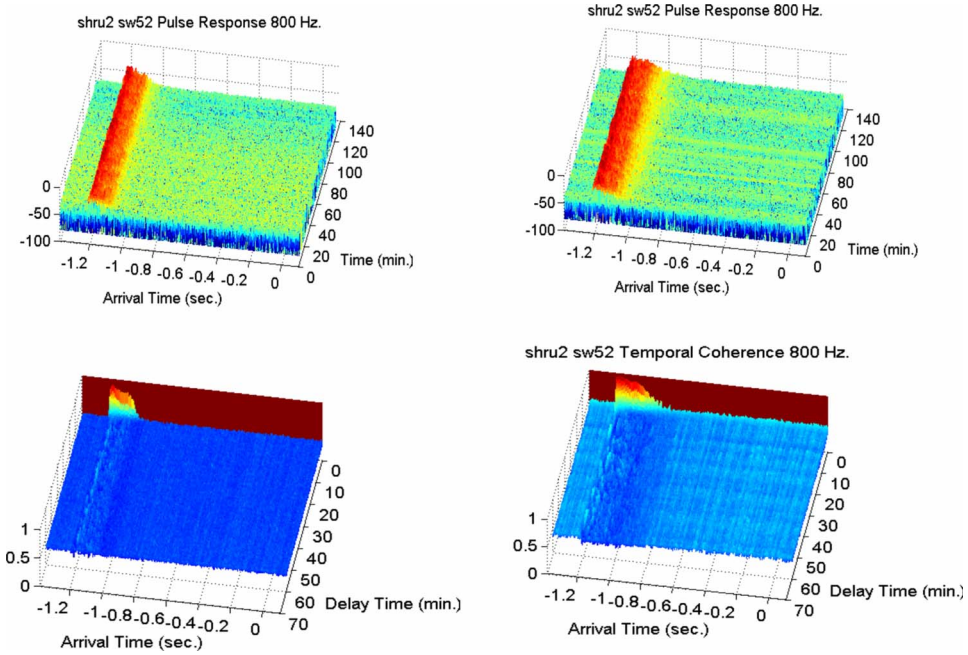


Fig. 4. (Color online) (Left top); Time history of the 800 Hz signal pulse responses at the time sample (arrow—left) of Fig. 2. (Left bottom) Coherency. (Right) Time history of 800 Hz signal pulse responses for sample time (arrow–right) of Fig. 2. (Right bottom) Coherency.

Neither mechanism seems to be at play for the 100 Hz transmission during the period of low internal wave potential energy [Fig. 3 (right)]. The modes are clean and distinct as is shown by the vertical array measurement and PE model comparison of Fig. 1. The model represents a smooth and homogeneous bottom and the remarkably close agreement with the modes observed with the SHARK Vertical Line Array (VLA) suggesting that the bottom appears to be smooth and reflective for the wavelength of the 100 Hz signal.

As a result, the coherence times are very long—at least 1 h during the low internal wave potential energy sample time. The coherence calculation cycles at a different rate for the successively higher modes. The first mode arrival is coherent for 1 h—basically unchanged in phase, amplitude, and wave form. The second mode is also perfectly coherent in time but undergoes a slow half-cycle oscillation during 1 h. This is a result of the phase cancellation in the coherence calculation; coherence values change slowly from one to zero over the 1 h period. However, the amplitude and wave form remain unchanged. Likewise, the third mode arrivals are coherent over the full hour but with a slow one cycle oscillation causing a coherence to vary from one to zero and then back one at the end of the hour. These observations cannot be explained by slow changes in sound speed since the first mode arrival time remains unchanged.

One possible cause is the tide. A change in depth of the propagation channel will not noticeably affect the fundamental mode which propagates directly down the channel. But, the second and higher modes have associated increasing angle of travel relative to the axis of the channel and will see a longer propagation path with an increase in channel depth. To investigate this possibility, the PE model was use to compute the phase difference for the modes of Fig. 1 resulting from an increase in channel depth of 0.3 m. The phase difference amongst modes rolls through about 360° for the first three mode arrivals, in close agreement with observations.

The coherency in the presence of high internal wave energy is quite different [Fig. 3 (left)]. The coherence for all modes falls to below 50% in less than 5 min, with the first mode having about the same decorrelation time as the higher modes. The decorrelation implied a change in wave form and not just a phase shift as before. Since the bottom appears flat, the mode distortion must come about from the internal wave scattering mechanism described by Duda and Preisig (1999) and Preisig and Duda (1997).

The same set of calculations for the 800 Hz signals shows dramatically different results as shown in Fig. 4. First, even during the low internal wave energy period there is no identifiable pulse arrival structure in time. The arrivals appear granular and unresolved in arrival time and inconsistent from one frame to the next. Without internal wave activity bottom scattering is likely the cause. A near continuum of mode arrivals results—probably from complicated scattering all along the path of propagation. The model interference patterns change rapidly with the slightest fluctuations in sound speed so the signal coherency falls off rapidly in a minute or two. Early arrivals, associated with lower order modes, are more coherent than later arrivals.

Similar data were taken at 200 and 400 Hz, but are not included here owing to page constraints. The striking changes in variability between the 100 and 800 Hz data can be seen as a smooth transition if the 200 and 400 Hz data are included

7. Conclusions

A consistent picture emerges when the above-presented results are combined with those of a previous paper that used the same method (DeFerrari, 2008). At very low frequencies the bottom appears smooth. Distinct modes are formed that are stable in time for 1 h or more. Equally long coherence times are observed for all higher order modes once the effects of tides are backed out. In the presence of internal waves, coherence times are reduced by a factor of 5, possibly owing in part to mode scattering by the sound speed fluctuations along the path of propagation. Still, all modes are seen to have equally long coherence times, suggesting that mode angle is not a significant factor for internal wave scattering.

As frequency increases, bottom scattering comes into play. Modes become less stable and spread in time and have shorter coherence times. Since higher order modes interact with the

bottom at steeper angle, they incur more scattering and have less coherence than lower modes. This is a consistent finding for the intermediate frequencies at both study sites.

At higher frequencies, mode scattering from the bottom dominates and a near continuum of short-lived, unstable modes replace the persistent arrivals of lower frequency. For still higher frequencies, coherence times are shorter still and at some point the *m*-sequences period exceed the coherence time so that signal processing gain is reduced and the signal is no longer detectable. The beginning of this effect is seen to have started for the 800 Hz data, where the pulse responses have reduced intensity during the high internal wave sample period.

Acknowledgments

The authors acknowledge Nick Shay, University of Miami, for helpful discussions on internal wave energy estimations. Work sponsored by ONR.

References and links

Collis, J. M., Duda, T. F., Lynch, J. F., and DeFerrari, H. (**2008**). "Observed limiting cases of horizontal coherence and array performance in a time-varying internal wave field," J. Acoust. Soc. Am. **124**, EL97–EL103.
DeFerrari, H. A. (**2008**). "Observations of low-frequency temporal and spatial coherence in shallow water," J. Acoust. Soc. Am. **124**, in press.
Duda, T. F., and Preisig, J. C. (**1999**). "A modeling study of acoustic propagation through moving shallow water solitary wave packets," IEEE J. Ocean. Eng. **24**, 16–32.
Evans, R. B. (**1983**) "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," J. Acoust. Soc. Am. **74**, 188–195.
Gill, A. E. (**1982**). *Atmosphere-Ocean Dynamics* (Academic, London) 661 pp.
McDaniel, S. T. (2006). "Mode coupling due to interaction with the seabed," J. Acoust. Soc. Am. **72**, 916–923.
Nguyen, B. H., DeFerrari, H. A., and Williams, N. (**1996**). "General purpose autonomous transmitter and receiver system for underwater acoustic experiments." IEEE J. Ocean. Eng. **21**, 235–244.
Preisig, J. C., and Duda, T. F. (**1997**). "Coupled acoustic mode propagation through continental shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269.
Yang, T. C. (**2006**). "Measurements of temporal coherence of sound transmissions through shallow water," J. Acoust. Soc. Am. **120**, 2595–2614.

Transmission loss measurements and geoacoustic sensitivity modeling at 1.2 kHz

Sean P. Pecknold, Kiyoshi W. Masui, and Paul C. Hines
Defence Research and Development—Atlantic, 9 Grove St., Dartmouth, Nova Scotia, Canada, B2Y 3Z7
sean.pecknold@drdc-rddc.gc.ca, kiyoshi.masui@drdc-rddc.gc.ca, paul.hines@drdc-rddc.gc.ca

Abstract: The Shallow Water Experiment 2006 was conducted off the coast of New Jersey in the summer of 2006. Defence Research and Development Canada—Atlantic performed a series of experiments designed to validate the use of rapid environmental assessment tools and methods to improve active sonar performance predictions. The sensitivity of acoustic propagation to a varying or uncertain environment is determined by examining the relative change of acoustic pressure caused by environmental variability, using the method described recently [Dosso *et al.*, J. Acoust. Soc. Am. **121**, 42 (2007)]. The variability of the modeled environmental parameters is based on measured and estimated oceanographic and geoacoustic properties. The resulting sensitivity is compared to measured transmission loss data at 1.2 kHz.

PACS numbers: 43.30.Pc [WC]
Date Received: May 12, 2008 **Date Accepted:** June 10, 2008

1. Introduction

The sensitivity of acoustic propagation to uncertain or variable environmental conditions, and the effect of this sensitivity on sonar performance predictions, has been the subject of much theoretical and experimental study.[1–4] During the Shallow Water Acoustics Experiment 2006 (SW06), a large-scale multilateral trial extending through July, August, and September of 2006, an examination of propagation sensitivity was undertaken in the context of rapid environmental assessment (REA), the salient questions being what the key sensitivity parameters are, and at what resolution they are required. The experiment took place at the Strataform East site, a natural laboratory established by the Office of Naval Research, off the coast of New Jersey.[5] A previous sensitivity study[3] in this area determined uncertainty in propagation loss due to internal waves as about 2 dB at 4 kHz. In this paper, one-way transmission loss measurements along several paths at a frequency of 1.2 kHz are shown, along with a characterization of the environment made using several REA techniques. This environmental information is used to examine the sensitivity of propagation in the experimental area to uncertainty or variability in environmental parameters, using a recently published full-field stochastic sensitivity measure.[6] This sensitivity is used to model mean expected transmission loss with error bounds on the model. The model results are then compared to the measured transmission loss along a line of approximately constant bathymetry parallel to the shelf break.

2. Transmission loss experiment and environment

The transmission loss experiment was performed August 1, 2006. The goal of the experiment was to measure acoustic propagation perpendicular to the continental shelf and parallel to the continental shelf along tracks for which the seabed and water column properties are accurately and contemporaneously measured. The research vessel *CFAV Quest* towed a dual-free-flooding ring acoustic projector along two tracks, at a speed of approximately 5 knots. Track 1 was parallel to the shelf, along the line from 39.02° N, 73.033° W to 39.223° N, 72.906° W. The water depth was nearly constant, at approximately 81 m. Track 2 was upslope, perpendicular to the first track. The projector was at a depth of approximately 50 m. The Track 1 data presented here is from 0.5 s linear frequency modulated pings (LFMs) with a center frequency of 1.2 kHz, with 200 Hz bandwidth, and with a 20 s repeat interval. The pings were transmitted

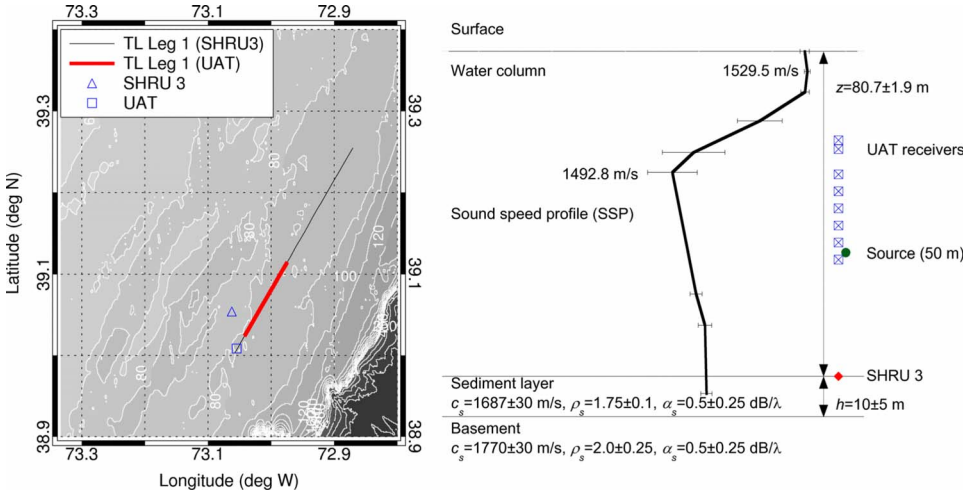


Fig. 1. (Color online) (Left) Configuration for Track 1 of the transmission loss experiment, on 1 August, 2006. (Right) Environmental model used for the sensitivity analysis.

with a source level of 200 dB re 1 μPa@1 m. These pings were recorded on the underwater acoustic target (UAT), a moored recorder with eight hydrophones, at depths of approximately 22, 24, 31, 35, 39, 43, 48, and 52 m. Data were also recorded on a bottom-moored single hydrophone recording unit (SHRU) operated by the Woods Hole Oceanographic Institute. The experimental configuration for Track 1 is shown in Fig. 1 (left).

3. Sensitivity measure and results

Considering environmental uncertainty or variability as a set of normally distributed model parameters, an ensemble of acoustic fields can be modeled by randomly perturbing these parameters. A stochastic sensitivity measure can then be defined by[6]

$$S = \frac{[\langle |\delta p|^2 \rangle]^{1/2}}{|p|}, \tag{1}$$

where p is an acoustic pressure field, δp the modeled perturbed field and $\langle . \rangle$ represents an ensemble average over the set of acoustic fields. This sensitivity measure can be calculated for a given propagation and environmental model to obtain a full range-depth sensitivity field.

In order to calculate this sensitivity, a set of environmental model parameters and a forward propagation model are required. The environmental model used for this sensitivity analysis is shown in Fig. 1 (right). Thirteen expendable sound velocity probes were dropped during the course of the experiment. Nine representative sound speed profile (SSP) depths were selected, and the mean and standard deviations of the sound speed at these points was determined. The value of 80.7 ± 1.9 m for the water depth is the mean and standard deviation measured using an echo sounder along the transmission loss run. The mean values and standard deviations for sediment layer and basement compressional sound speed, density, and compressional attenuation, and for sediment layer thickness, were determined using a genetic algorithm inversion routine,[7] using the OASES transmission loss (OAST)[8] wave number integration transmission loss model as the forward model. The sea bottom was assumed to be a fluid sediment layer atop a fluid basement, and the mean sound speed profile and depth were used. The inversion inputs were sets of transmission loss data from the continuous wave transmissions at 1.2 kHz, for the bottom six receivers of the UAT. The values found for the parameters are consistent with previous measurements in the area.[9] The parameters were also compared to the results obtained from "pseudo cores" made using a free-falling cone penetrometer (FFCPt).[10] Here, dynamic pore pressure and acceleration are measured as a function of penetration into the

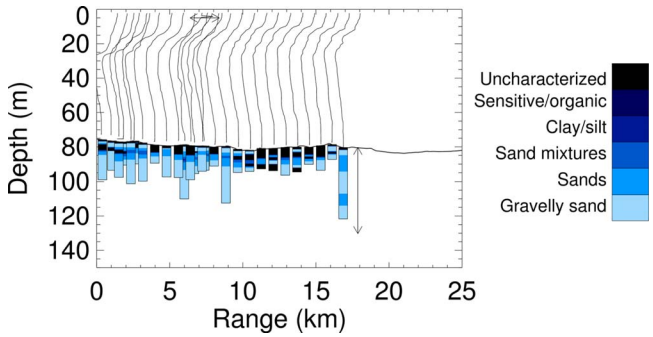


Fig. 2. (Color online) FFCPt drops and sound speed profiles taken along propagation run 1, on August 1, 2006. Horizontal arrow shows scale of sound speed and has an extent of 50 m/s. Vertical arrow shows scale of FFCPt penetration depth and has depth of 0.5 m.

seabed. These values are used to calculate shear strength, which is then translated into a sediment type using the method of Robertson.[11] Figure 2 shows the results of the FFCPt drops together with sound speed profiles measured along the transmission loss run. The upper part of the sediment layer is primarily sand and gravelly sand, with some sand-silt mixtures. Wind speed averaged approximately 12 knots through the course of the experiment, and was held constant in the model. The PECan parabolic equation model[12] was used as the forward propagation model.

Full-field sensitivity measures for the acoustic pressure field [Eq. (1)] were generated by randomly perturbing the following sets of parameters: The total water column SSP (all nine points perturbed independently); each of the seven geoacoustic parameters; the water depth; and the set of all model parameters. Five hundred perturbed fields were generated for each of these ten cases. In general, sensitivity is greatest for SSP, followed by water depth, and then sediment attenuation, sound speed, and density. The acoustic field at 1.2 kHz was insensitive to perturbations in sediment thickness and the properties of the basement; in essence, the sediment layer acted as a half space. Three examples are shown in Fig. 3, sensitivity to perturbations of: sediment compressional sound speed; SSP; and all model parameters. The sensitivity is both range and depth dependent. Note that the increased sensitivity to the sediment layer properties (all of which behave in a similar manner to the sound speed) at shallower depths at long ranges is due to the propagation path—the downward refracting profile forces interaction with seabed. Sensitivity to the water column SSP is greatest near 30 and 50 m. Fluctuations in the SSP near the minimum can cause an acoustic duct to form around 30 m, or create a nearly isovelocity profile below the thermocline, leaving more acoustic energy at the source depth. The overall sensitivity shows the impact of fluctuations in water depth, with thin bands of higher sensitivity above and below the black line that denotes the seabed.

4. Transmission loss and model comparison

The transmission loss (TL) along Track 1, measured for two of the eight UAT receiver depths (22 and 52 m) and for the SHRU 3 receiver, is shown in Fig. 4. TL was calculated for the matched filter output of the 1.2 kHz LFM pings, using a replica that was Doppler shifted to correct for tow speed. The modeled coherent TL at these depths using the environmental model mean values is also shown in the top row, while the top and middle rows show the mean of the perturbed model runs with $\pm 1\sigma$ values. The mean and standard deviation for the modeled transmission loss is calculated using the sensitivity ensembles previously described, but with the mean and standard deviation calculations done on the transmission loss in dB, to describe the statistics of the model output rather than the acoustic field. The top row shows model outputs where all parameters are perturbed, while the middle row shows those with only the sound speed profile in the water column perturbed. The sensitivities to the sediment parameters, water depth, and sound speed profile with respect to range are shown in the bottom row. The sensitivity to

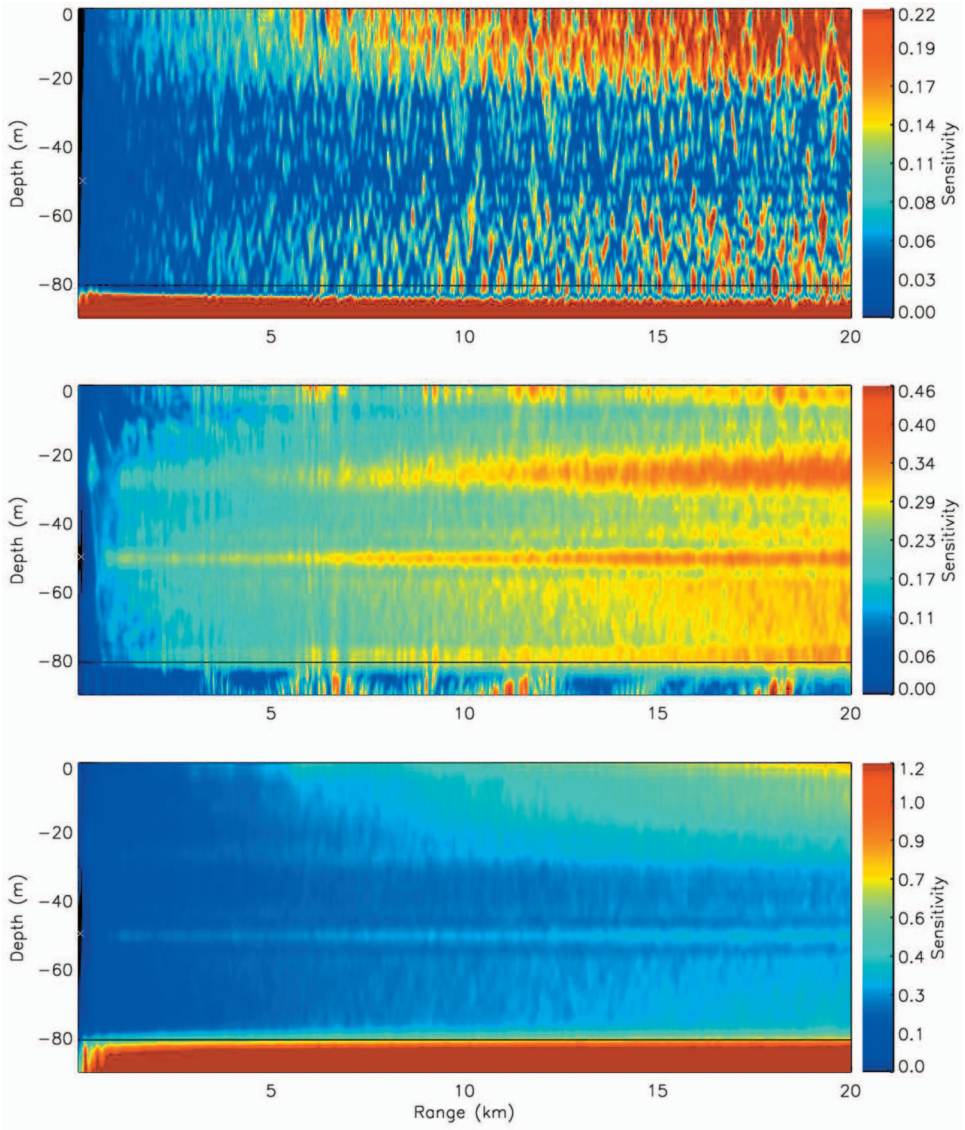


Fig. 3. Sensitivity of acoustic pressure field to: (top) sediment compressional sound speed; (middle) water column sound speed profile; (bottom) all model parameters. Source depth is indicated with an "X." Note that the scales differ.

sediment attenuation was higher than expected, due to the high standard deviation on the value, i.e., the typical perturbation used in modeling was large relative to the value of the parameter. Improved estimates of the geoacoustic parameters will most likely result in a decreased sensitivity measure with respect to those parameters. The model results are averaged over 100 m in range, or 20 grid points.

The one-sigma model results with all parameters perturbed bracket between 86.9% and 96.5% of the measured transmission loss points. This indicates that the specified uncertainty on the model parameters is too large. The sensitivity values in Fig. 3 show that the model variability is driven primarily by SSP and sediment attenuation for the upper- and mid-level receivers, and by water depth for the bottom receiver. Constraining the perturbations to the SSP

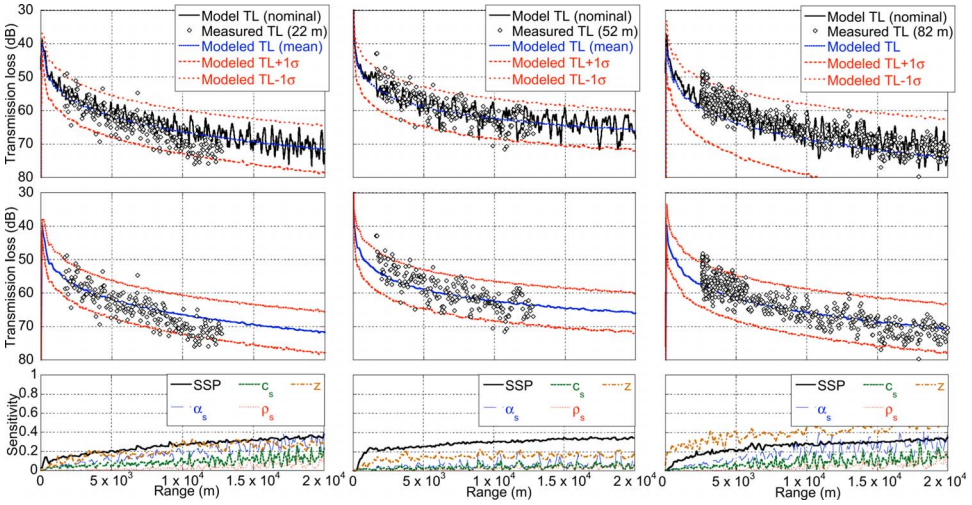


Fig. 4. (Color online) Top row: Measured and modeled transmission losses for 22, 52, and 82 m depth, with mean and standard deviation of model ensembles for perturbation of all model parameters. Middle row: Measured transmission losses for 22, 52, and 82 m depth, with mean and standard deviation of model ensembles for perturbation of sound speed profile. Bottom row: Sensitivities to sediment and water column properties for 22, 52, and 82 m depths.

has little to no impact on the model variability at 52 m depth (and on the other receivers between 30 and 50 m). The variability at 22 m is reduced somewhat, decreasing the included measured points from 91.6% to 87.8%. The model variability at the bottom is considerably reduced by constraining the perturbed variables to just the SSP, although over 95% of the data points are still bracketed. Even with this constraint, the uncertainty in modeled propagation is between 10 and 12 dB for ranges from 2 to 20 km, considerably greater than that found previously.[3] As the propagation variability is in general driven by the variability in the sound speed profile, improved measurements of the sediment layer properties would not necessarily result in improvements in accuracy of propagation modeling.

5. Conclusions

The sensitivity of the acoustic field along a track parallel to the shelf slope at the SW06 experimental site was modeled at 1.2 kHz, and was found to depend on sound speed profile and then sediment attenuation, sound speed, and density, with water depth playing a greater role for areas near the bottom. Little dependence was found on either sediment thickness or basement properties. A comparison of these results with measured transmission loss data suggests that the environmental variability measured sufficed to explain the variability in transmission loss data, and resulted in model variability beyond that observed in the measured data. In fact, the variability in sound speed profile alone was able to explain observed transmission loss variability, therefore further constraints on the bottom would not greatly improve model prediction capabilities unless water column observations are also improved.

Acknowledgments

We wish to acknowledge Arthur Newhall and the Woods Hole Oceanographic Institute for the provision of SHRU data used in the analysis, and the Office of Naval Research for their support in organizing the Shallow Water Acoustics Experiment 2006.

References and links

[1]L. Sha and L. W. Nolte, "Effects of environmental uncertainties on sonar detection performance prediction," J. Acoust. Soc. Am. **117**, 1942–1953 (2004).
[2]R. T. Kessel, "A mode-based measure of field sensitivity to geoacoustic parameters in weakly range-dependent environments," J. Acoust. Soc. Am. **105**, 122–123 (1999).

[3]K. D. LePage, C. Holland, and J. A. Goff, "Transfer of oceanographic and bottom variability into shallow water propagation and reverberation uncertainty," J. Acoust. Soc. Am. **114**, 2311–2312 (2003).
[4]N. G. Pace and F. B. Jensen, *Impact of Littoral Environmental Variability on Acoustic Predictions and Sonar Performance* (Kluwer, Dordrecht, 2002).
[5]J. A. Goff, D. J. Swift, C. S. Duncan, L. A. Mayer, and J. Hughes-Clarke, "High–resolution swath sonar investigation of sand ridge, dune and ribbon morphology in the offshore environment of the New Jersey Margin," Mar. Geol. **161**, 307–337 (1999).
[6]S. E. Dosso, P. M. Giles, G. H. Brooke, D. F. McCammon, S. P. Pecknold, and P. C. Hines, "Linear and nonlinear measures of ocean acoustic environmental sensitivity," J. Acoust. Soc. Am. **121**, 42–45 (2007).
[7]P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions," J. Acoust. Soc. Am. **95**, 770–782 (1994).
[8]H. Schmidt, *OASES Version 1.6: Application and Upgrade Notes.* (Massachusetts Institute of Technology, Cambridge, MA, 1993).
[9]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[10]J. Osler, A. Furlong, H. Christian, and M. Lamplugh, "The integration of the Free Fall Cone Penetrometer (FFCPT) with the Moving Vessel Profiler (MVP) for the rapid assessment of seabed characteristics," Int. Hydrogr. Rev **7**, 45–53 (2006).
[11]P. K. Robertson, "Soil classification by the cone penetration test," Can. Geotech. J. **27**, 151–158 (1990).
[12]G. H. Brooke, D. J. Thomson, and G. R. Ebbeson, "PECan: A Canadian parabolic equation model for underwater sound propagation," J. Comput. Acoust. **9**, 69–100 (2001).

Direct measurement of sediment sound speed in Shallow Water '06

Jie Yang, Dajun Tang, and Kevin L. Williams
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698
jieyang@apl.washington.edu, djtang@apl.washington.edu, williams@apl.washington.edu

Abstract: Knowledge of sediment sound speed is crucial for predicting sound propagation. During the Shallow Water '06 experiment, *in situ* sediment sound speed was measured using the Sediment Acoustic-speed Measurement System (SAMS). SAMS consists of ten fixed sources and one receiver that can reach a maximal sediment depth of 3 m. Measurements were made in the frequency range 2–35 kHz. Signal arrival times and propagation distances were recorded, from which sediment sound speed was determined. Preliminary results from three deployments show that SAMS was capable of determining sediment sound speed with uncertainties less than 1.6%. Little dispersion in sediment sound speed was observed.

PACS numbers: 43.30.Pc, 43.30.Ma, 43.30.Xm [WC]
Date Received: March 26, 2008 **Date Accepted:** May 19, 2008

1. Introduction

Acoustic interaction with the sea bottom is, in many cases, an essential component of sound propagation in a shallow water waveguide. *In situ* geoacoustic properties of the seabed, however, are difficult to obtain. Since the 1950s, *in situ* direct measurements have been carried out to study the sediment geotechnical properties.[1] More recently, sediment geoacoustic properties have been measured within the surfacial layer in the frequency range 1–100 kHz using either manually buried acoustic systems (Refs. 2–4) or specially designed underwater mechanical systems. Refs. 5–8, from the latter category, present systems that can penetrate into the sediment through their own gravitational forces. The *In Situ* Sediment geoacoustic Measurement System (Refs. 5–7) is designed for *in situ* measurements within the topmost 30 cm while the acoustic lance (Ref. 8) has a maximum penetration depth of 5 m.

As part of the experimental effort in Shallow Water '06 (SW06), the Sediment Acoustic-speed Measurement System (SAMS) was used to directly measure the sediment sound speed. SAMS is driven into the seabed by a powerful vibrocore, which allows precise penetration depth up to 3 m with arbitrary step size. The ground truth measurements are valuable not only in studying *in situ* sediment properties but also in providing sediment geoacoustic data to which inversion results can be compared.

This paper is organized as follows. In Sec. 2, the analyses of calibration and sediment data are presented. Section 3 summarizes and discusses future directions. System uncertainty analysis is given in the Appendix.

2. Data analysis

SAMS was deployed during the SW06 field experiment. Four data sets, one in the water column as calibration and three in the sediment, were recorded. For all data sets, three linear-frequency-modulated (LFM) "chirps" were used in the frequency bands 2–11, 10–21, and 20–35 kHz, which are referred to as low-, mid-, and high-frequency (LF, MF, and HF) in later analysis.

2.1 Analysis of calibration data

Calibration data were acquired in a bay environment and relatively far from the positions of the sediment measurements. For calibration, SAMS was deployed such that both sources and re-

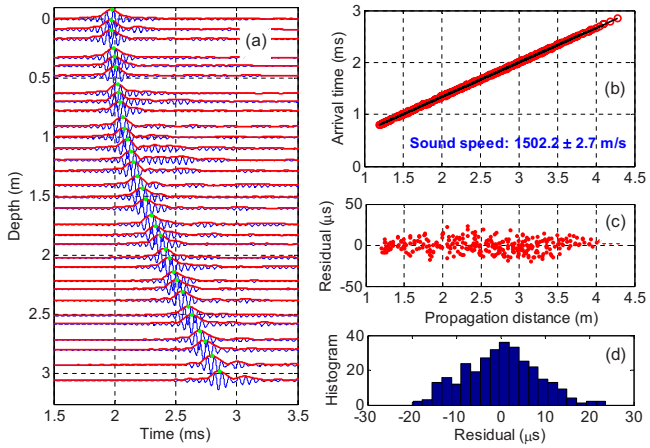


Fig. 1. (Color online) Pulse compression results at MF. (a) Waveforms received from source 10; (b) linear regression results; (c) linear regression residual scatter plot; (d) residual histogram.

ceiver were always within the water column. The analysis of the calibration data is used to augment the system geometry measurements and establish system uncertainty.

SAMS consists of a 3-m-tall triangular frame and a 2-m-long extension beam. Three sources are fixed on the triangular frame with the other seven on the extension beam. The receiver is located in the center of the frame and can be driven vertically into the sediment using a vibrocore. Horizontal distances between sources and receiver are between 1.18 and 2.97 m.

Data were taken at 31 receiver depths with a stepsize of about 0.1 m. At each depth, the ten sources sequentially transmitted three chirps and each was repeated five times. Pulse compression was carried out as an initial step in processing the data. In Fig. 1(a), pulse compressed waveforms received from source ten are plotted at their corresponding receiver depths. Signal arrival time is defined as the time at the peak of the envelope and highlighted with a dot. With signal arrival times defined, the speed of sound in water can be determined from the linear regression of the arrival times and distances between the sources and receiver. Five repeated pings are used to find the averaged arrival times yielding a total of 310 data points. The linear regression result, Fig. 1(b), is 1502.2 m/s with 2.7 m/s uncertainty at the 95% confidence level. The calculation of uncertainty assumes that the residuals (difference between data and fit function) are random and follow a normal distribution of zero mean and constant variance. A scatter plot of the residuals is shown in Fig. 1(c). Both the scatter plot and its histogram [Fig. 1(d)] indicate that the distribution of the residual is close to normal. Similar procedures are repeated for the LF and HF calibration data. The curve fitting results for the speed of sound in water are 1503.4 ± 6.7 and 1503.1 ± 3.4 m/s, respectively.

The sound speeds determined are close to each other with the confidence interval at MF and HF almost completely enclosed by that of the LF. The higher uncertainty at LF is due to a roughly 58% decrease in the total number of data points included in the curve fitting process. Beyond 2 m, the LF calibration data showed signals, possibly due to the tube waves, arriving prior to the direct arrivals and, therefore, were excluded from the curve fitting. Unfortunately, there was no conductivity-temperature-depth (CTD) record when the calibration data were taken. Concurrent ship data only provide temperature and salinity at the sea surface. Therefore, a total of 120 historical summer CTD data around the area were sought for reference. Individual distance to the SAMS calibration position varied from 16 to 52 km and water depth ranged 13–40 m. A strong thermocline was observed starting at around 10 m and water sound speed varied from 1520 m/s at the sea surface to 1495 m/s at 35 m. For calibration, bathymetry showed a 23 m water depth with SAMS suspended 5–6 m below the surface, i.e., measurements were taken at 6–9 m in depth. Compared with historical CTD data, the calibration results, 1503.4 ± 6.7, 1502.2 ± 2.7, and 1503.1 ± 3.4 m/s, respectively, are reasonably within the

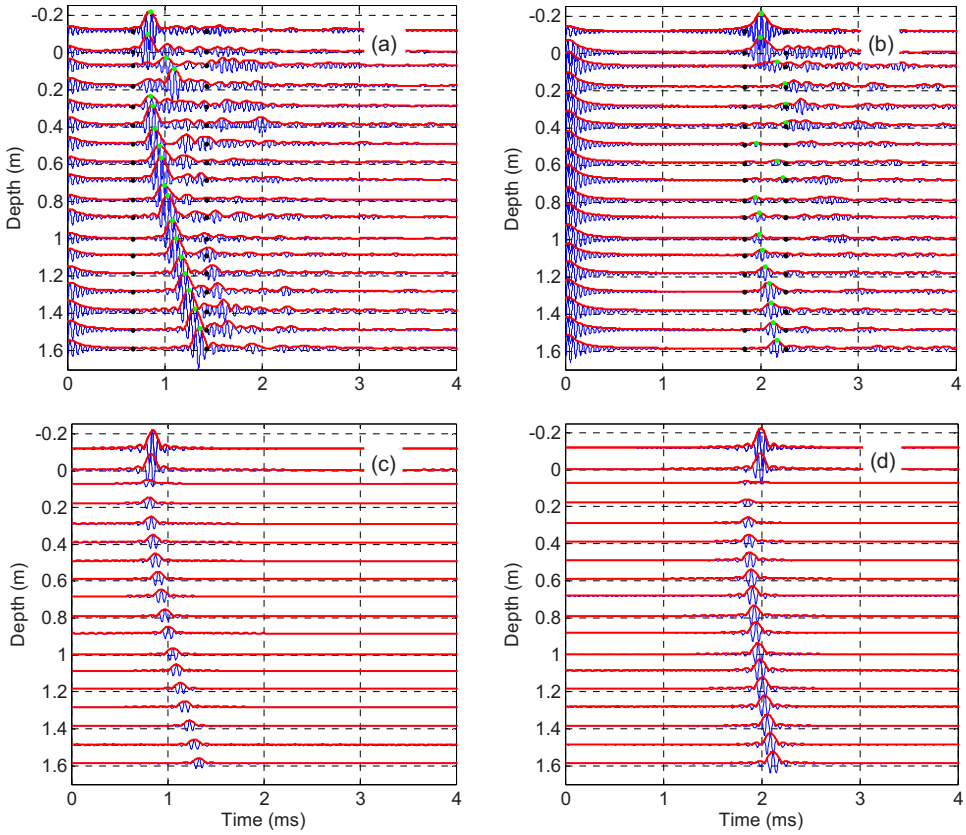


Fig. 2. (Color online) MF Pulse-compressed waveforms from: (a) source 1; (b) source 10 at different receiver depths. (c) and (d) are half-space Green's function simulations of (a) and (b), respectively.

variation range. More importantly, the acoustic data are of good quality and repeatable at different frequency bands, supporting the conclusion that the water sound speeds measured by SAMS are close to the true values at the time.

2.2 Analysis of sediment data

To start, two cases in the MF band from position 2 are chosen to show the characteristics of sediment data. The first case, Fig. 2(a), draws the signals from source 1 at 19 receiver depths with the maximum penetration depth around 1.6 m. Data are processed in the same way as the calibration. The receiver starts at about 10 cm above the sediment surface, which makes the first depth sample less than zero in the figure. In determining the signal arrival time, a time window is specified (bounded by the black dots) within which the peak of the signal envelope is recognized and highlighted with a green dot. It is quite obvious that there are two erroneous readings of the peak time in Fig. 2(a) as the receiver first enters the sediment. For sources that are further away from the receiver, there are more such occurrences as shown in Fig. 2(b). Data of this kind are carefully excluded from analysis. Signal-to-noise ratio drops considerably for the geometries realized in Fig. 2(b) due to the combined effects of longer acoustic path through the sediment and ray bending by the critical angle. A two-half space Green's function is used to simulate the scenario of Figs. 2(a) and 2(b) and the results are shown in Figs. 2(c) and 2(d). The critical angle effect is quite apparent in Fig. 2(d).

Table 1. Summary of sediment sound speed results.

		C_b (m/s)
Position 1	LF	1614.8 ± 8.7
	MF	1622.1 ± 12.5
Position 2	LF	1597.7 ± 11.0
	MF	1598.6 ± 9.8
Position 3	LF	1588.2 ± 15.8
	MF	1611.6 ± 24.8

To find sediment sound speed, the water sound speed close to the water–sediment interface is required. Throughout the experiment, CTD records show a very stable water sound speed at depths beyond 70 m. Based on CTD records, the water sound speed is 1496 m/s. Sediment sound speed is assumed homogeneous within the penetrated depth, 1.6 m. Ray tracing is carried out for each data point by varying sediment sound speed in the model and the closest match of arrival times between measurement and ray tracing determines the *in situ* sediment sound speed. The uncertainties are determined in the same manner as for the calibration (Sec. 2.1). Assuming uncertainty comes entirely from propagation in the sediment, the time and distance that are spent in water are removed from the total time and distance. The uncertainty in sediment sound speed is then calculated using the residuals at 95% confidence level.

Results are summarized in Table 1 and Fig. 3. Results from positions 1 and 2 show similar uncertainty bounds around 10 m/s and little dispersion between the two frequency bands. There is about a 20-m/s sound speed difference between the two positions, which is believed to be the true spatial variation. Results at position 3 have much larger uncertainties than the other two. Signs of signal degradation were observed (not shown here), which results in a 50% higher mean residual than the other two positions. Geological studies around the SW06 region (Ref. 9), using interpreted chirp seismic reflection data, indicate a substantial difference in sediment properties at position 3 from positions 1 and 2, which is believed to be the cause for higher uncertainty at position 3. In addition, acoustic measurements of bottom reflection (Ref. 10) were made in the vicinity of SAMS positions 1 and 2 in the frequency range 1–20 kHz. The geoacoustic inversion results of sediment sound speed, with co-located coring and stratigraphic studies, are consistent with the direct measurement results using SAMS.

3. Summary and future directions

In this paper, direct measurements of sediment sound speed using the Sediment Acoustic-speed Measurement System (SAMS) have been presented. The calibration data were first analyzed to establish the system uncertainty, which is approximately 3 m/s. Sediment data were taken at three positions. Sediment sound speeds and uncertainties are summarized in Table 1 and Fig. 3. Results indicate a 20-m/s sound speed variation between positions 1 and 2. At position 3, the

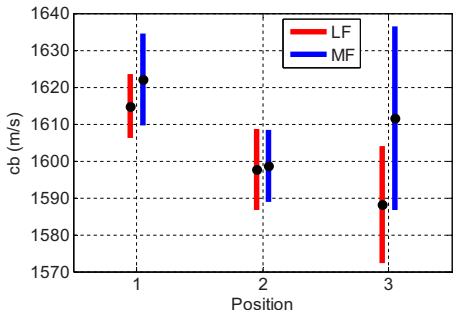


Fig. 3. (Color online) Sediment sound speeds with uncertainties at three positions.

increase in uncertainty may be attributed to the sediment properties based on geological studies around the central experimental area in SW06. The sediment sound speeds found at positions 1, 2, and 3 are 1618 ± 11, 1598 ± 10, and 1600 ± 20 m/s, respectively. Little dispersion in sediment sound speed was observed. Direct measurement of sediment sound speed dispersion has been found in sandy sediments (Refs. 2 and 11). The dispersion was observed to be at its greatest in the frequency range 800–2000 Hz. In this work, the frequency coverage is higher than the pronounced transition region of dispersion, which may explain the observed lack of dispersion. Future directions include improvement of system uncertainty in sediment sound speed, determination of sediment attenuation and its dispersion relation, and depth dependence of sediment geoacoustic properties.

Acknowledgment

This work was supported by the Office of Naval Research.

Appendix: System uncertainty analysis

For a system like the SAMS, the dimension and propagation time uncertainties limit the resolution of sound speed measurement. Uncertainty can come from both measurements and methodology utilized to analyze data. In this work, part of the uncertainty comes from measurement of distances, i.e., horizontal distance and initial depth offset between sources and receiver; initial depth offset between receiver and the sediment surface; and receiver depth reading. The reading of the arrival times falls in the latter category.

Following Ref. 12, a general function q with multiple variables $(x, y, z, \ldots)$ has uncertainty:

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \left(\frac{\partial q}{\partial y}\delta y\right)^2 + \left(\frac{\partial q}{\partial z}\delta z\right)^2 + \cdots}. \tag{A1}$$

In Eq. (A1), variables $x, y, z, \ldots$ are independent measurements with uncertainties $\delta x, \delta y, \delta z, \ldots$. For this work, the general function is the speed of sound in water, c_w, which is the ratio of distance and time:

$$c_w = \frac{r(x,y,z,\cdots)}{t} = \frac{\sqrt{x^2 + (h+d)^2}}{t}, \tag{A2}$$

where r is the slant distance, t is travel time; x, h, and d are horizontal distance, receiver depth, and initial vertical distance between source and receiver. Following Eq. (A1), the uncertainty in calibration can be written as

$$\delta c_w = \sqrt{\left(\frac{\partial c_w}{\partial x}\delta x\right)^2 + \left(\frac{\partial c_w}{\partial h}\delta h\right)^2 + \left(\frac{\partial c_w}{\partial d}\delta d\right)^2 + \left(\frac{\partial c_w}{\partial t}\delta t\right)^2}. \tag{A3}$$

The individual uncertainties $\delta x, \delta h, \delta d$ are defined as 0.9, 0.1, and 0.5 cm. Specifically, δx is set to $\frac{1}{4}$ of the source dimension; δd is set to $\frac{1}{4}$ of the receiver dimension; δh accounts for depth reading uncertainty. In Eq. (A3), the most difficult part is to determine δt. The receiver is embedded inside a stainless steel tube with two rectangular windows open on the side. The combination of direct arrival and reflections off the window may slightly change the location of the signal peak. Assuming this window effect is random, δt can be determined using the mean residual obtained from Fig. 1 as 7.5 μs. Figure 4 shows the uncertainty corresponding to each source. For each of them, as depth increases, the uncertainty decreases from top to bottom. The black dashed line is the mean value for each source. It is obvious that sources that are closer to the receiver have higher uncertainty and spreading. The least uncertainty for an individual measurement is around 5 m/s.

In calibration analysis, data recorded from all ten sources were used in the linear regression to find the speed of sound in water, i.e., a relationship between individual (Fig. 4) and overall uncertainty (Fig. 1) needs to be clarified. The overall system uncertainty can be deter-

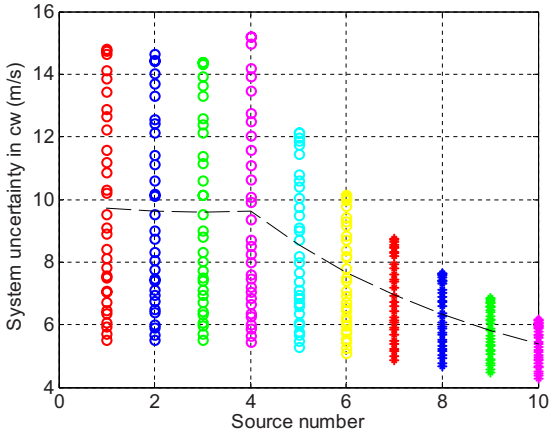


Fig. 4. (Color online) System uncertainty in the speed of sound in water.

mined from the individual uncertainties in Fig. 4 as a two-step process. First, assume an ideal system with no uncertainty, i.e., find propagation times by dividing measured propagation distances by a fixed sound speed. The value c_w=1502.2 m/s, determined from calibration at MF, is used. The linear regression shows a perfect fit between propagation distance and time with zero uncertainty. Second, convert the maximum uncertainty of each source δc_w, as in Fig. 4, to its equivalent uncertainty in distance, i.e., by multiplying δc_w with corresponding propagation time t. Then, the measured propagation distances were added a random quantity in the range of $\pm t \times \delta c_w$. The linear regression is carried out again by forcing c_w=1502.2 m/s. The system uncertainty at 95% confidence interval is calculated in a similar fashion as in Sec. 2. This uncertainty, determined to be ±1.6 m/s, is the overall system uncertainty.

References and links

[1]E. L. Hamilton, G. Shumway, H. W. Menard, and C. J. Shipek, "Acoustic and other physical properties of shallow-water sediments off San Diego," J. Acoust. Soc. Am. **28**, 1–15 (1956).
[2]A. Turgut and T. Yamamoto, "Measurements of acoustic wave velocities and attenuation in marine sediments," J. Acoust. Soc. Am. **87**, 2376–2383 (1990).
[3]A. I. Best, Q. J. Huggett, and A. J. K. Harris, "Comparison of in situ and laboratory acoustic measurements on Lough Hyne marine sediments," J. Acoust. Soc. Am. **110**, 695–709 (2001).
[4]G. B. N. Robb, A. I. Best, J. K. Dix, P. R. White, T. G. Leighton, J. M. Bull, and A. Harris, "Measurements of the in situ compressional wave properties of marine sediments," IEEE J. Ocean. Eng. **32**, 484–496 (2007).
[5]A. Barbegelata, M. D. Richardson, B. Miaschi, E. Muzi, P. Guerrini, L. Troiano, and T. Akal, "ISSAMS: An in situ sediment acoustic measurement system," in *Shear Waves in Marine Sediments*, edited by J. M. Hoven, M. D. Richardson, and R. D. Stoll (Kluwer, Dordrecht, 1991).
[6]M. D. Richarson, D. L. Lavoie, and K. B. Briggs, "Geoacoustic and physical properties of carbonate sediments of the lower Florida Keys," Geo-Mar. Lett. **17**, 316–324 (1997).
[7]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).
[8]S. S. Fu, R. H. Wilkens, and L. N. Frazer, "Acoustic lance: New in situ seafloor velocity profiles," J. Acoust. Soc. Am. **99**, 234–242 (1996).
[9]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[10]J. W. Choi, P. H. Dahl, and J. A. Goff, "Observations of the R reflector and sediment interface reflection at the Shallow Water '06 Central Site," J. Acoust. Soc. Am. **124**, EL128–EL134 (2008).
[11]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).
[12]J. R. Taylor, *An Introduction to Error Analysis*, 2nd ed. (University Science Books, 1997).

In situ measurements of velocity dispersion and attenuation in New Jersey Shelf sediments

Altan Turgut
Naval Research Laboratory, Acoustics Division, Washington, DC 20375
turgut@nrl.navy.mil

Tokuo Yamamoto
University of Miami, RSMAS, Div. of Applied Marine Physics, Key Biscayne, Florida 33149
tyamamoto@rsmas.miami.edu

Abstract: The existence of acoustic velocity dispersion and frequency dependence of attenuation in marine sediments is investigated using *in situ* measurements from a wideband acoustic probe system during the Shallow Water 2006 experiment. Direct-path pulse propagation measurements show evidence of velocity dispersion within the 10–80 kHz frequency band at two silty-sand sites on the New Jersey Shelf. The measured attenuation in dB/m shows linear frequency dependency within the 10–80 kHz frequency band. The measured velocity dispersion and attenuation curves are in good agreement with those predicted by an extended Biot theory [Yamamoto and Turgut, J. Acoust. Soc. Am. **83**, 1744-1751 (1988)] for sediments with a distribution of pore sizes.

PACS numbers: 43.30.Ma, 43.20.Jr [WC]
Date Received: March 25, 2008 **Date Accepted:** June 19, 2008

1. Introduction

Compressional wave velocity and attenuation are two of the most important geoacoustic parameters that control sound propagation in shallow water. The existence of velocity dispersion and nonlinear frequency dependence of attenuation within the seabed might drastically change the predictions of propagation models that commonly use constant velocity and linear frequency dependence of attenuation. Several measurement techniques have been used to measure the compressional wave velocity and attenuation at different frequency bands in different types of marine sediments. Strong velocity dispersion in well-sorted beach sediments in the 1–30 kHz frequency band has been observed by using cross-probe measurements.[1] Critical-angle measurements of reflection coefficient[2] and shotgun/sediment-probe measurements[3] also indicated strong velocity dispersion in granular marine sediments. A lesser degree of velocity dispersion within the 25–100 kHz frequency region has been observed in medium-sand sediments in the Gulf of Mexico by using a sediment probe system.[4] More recent time-of-flight measurements by using buried source and receivers in the same area showed strong velocity dispersion within the 1–5 kHz region.[5] A reflection measurement technique[6] showed almost negligible velocity dispersion within the 0.1–10 kHz frequency region in silty-sand sediments of Malta Plateau and New Jersey Shelf. An earlier version of the wideband acoustic probe system[7] showed weak velocity dispersion in silty sediments and no velocity dispersion in muddy sediments within the 20–100 kHz frequency band. In this paper, wideband (10–80 kHz) measurements of velocity dispersion and attenuation in silty sand are reported.

In Sec. 2, a brief description of measurements and methods for measuring *in situ* velocity dispersion and attenuation are given. Then, measured velocity dispersion and attenuation are compared with those calculated from an extended Biot model.[8] In Sec. 3, the extended Biot model is briefly described and sound speed and attenuation predictions are presented for sediments having different distribution of pore sizes. Finally, in Sec. 4, summary and conclusions are given.

2. Experiment

As a part of sediment charcterization effort during the SW06 experiment, a wideband acoustic probe system was deployed at two different sites on the New Jersey Shelf. At Site-1 (39.0235N, 73.0348W), additional high-frequency scattering and propagation measurements were conducted. At Site-2 (39.00145N, 73.1202W), independent measurements of sound speed, density, porosity, and grain size were available from a 3-m-long sediment core. The acoustic probe system includes a self-contained data acquisition unit, four acoustic probes (one source and three receivers), and about 400 kg of lead weights. The self-contained data acquisition unit is programmed on board the research vessel based on a planned pulse transmission schedule. The system can be operated at water depths up to 2000 m with up to 2 m penetration into the sediment. Each probe is 20 mm in diameter and has an adjustable length. The acoustic transducers are cylindrical rings with OD=15 mm, ID=12 mm, and height=20 mm. Loading effects might be significantly different when the acoustic transducer is placed in hard (sandy) sediments than those in the water column. These differences might introduce measurement errors if the waveforms measured by a single receiver in the water and sediment. In our measurements, the wideband pulse signals received at two buried receivers are used for measuring velocity and attenuation as a function of frequency. Spectral ratio method[1] is used to calculate and phase velocity and attenuation from the measured pulse spectra. The phase delay between two receivers is calculated as

$$\Delta\phi(\omega) = \tan^{-1}\{\mathrm{Im}[S_2(\omega)/S_1(\omega)]/\mathrm{Re}[S_2(\omega)/S_1(\omega)]\}, \tag{1}$$

where $S_1(\omega)$ and $S_2(\omega)$ are spectra of the pulses received at distances d_1 and d_2, respectively. When the distance between two receivers is larger then the acoustic wavelengths, the above equation provides a wrapped phase delay in the range $[-\pi\ \pi]$. To remove the phase wrapping, a reference velocity c_0 is used to align the received pings by applying a constant phase shift to $S_2(\omega)$ as $S_2'(\omega)=S_2(\omega)\exp[-i\omega(d_2-d_1)/c_0]$. In Eq. (5), the spectrum $S_2(\omega)$ is replaced by $S_2'(\omega)$ and a new phase delay $\Delta\phi'$ is obtained. Then, the phase speed is calculated using the new phase delay as

$$c_p = c_0\left(1+\frac{c_0\Delta\phi'}{\omega(d_2-d_1)}\right)^{-1}. \tag{2}$$

The spectral ratio of two pulses received by two receivers is also used for measuring attenuation as a function of frequency. Spectral ratios are calculated both in the water column and sediment so that geometrical spreading and different receiver sensitivities can be taken into account automatically. The attenuation in dB/m is calculated from the spectral ratio as

$$\alpha = \frac{8.686}{(d_2-d_1)}\ln[|S_1(\omega)/S_2(\omega)|/|Q_1(\omega)/Q_2(\omega)|], \tag{3}$$

where $Q_1(\omega)$ and $Q_2(\omega)$ are spectra of the pulses transmitted in water and received at distances d_1 and d_2, respectively. Three Gaussian-windowed LFM pulses (5–25 kHz, 20–50 kHz, and 40–100 kHz) are transmitted by the source transducer, and measured pulse spectra $S_1(\omega)$ and $S_2(\omega)$ are used in Eqs. (1)–(3) to calculate phase velocity and attenuation. Figure 1 shows the measured phase velocity and attenuation at Site-1 and Site-2. Note that compressional wave velocity gradually increases for each frequency band from 10 to 80 kHz. The velocity dispersion is slightly stronger at the Site-1 especially within 30–40 kHz frequency band. The gaps in the measured data are due to the discarded low signal to noise ratio regions at both sides of the Gaussian-shaped spectra. An extended Biot theory, described in the next section, was used to calculate velocity and attenuation curves for silty sand with a log-normal pore-size distribution. The sediment physical properties, used in the calculations, are inferred from the core data and given in Table 1. The calculated phase velocity and attenuation curves are also plotted in Fig. 1 with pore-size standard deviations of σ=0.0, σ=1.25, and σ=2.0. The Biot theory with the

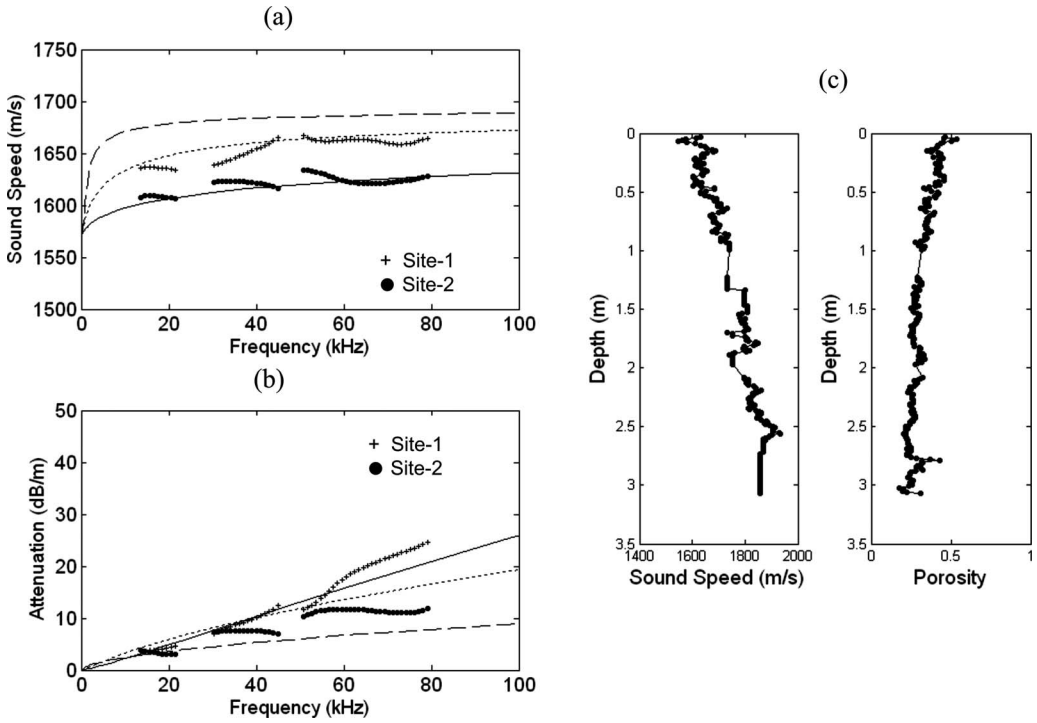


Fig. 1. (a-b) Comparison of measured velocity and attenuation curves with those calculated by using the extended Biot theory. The physical properties in Table I with pore-size standard deviations of $\sigma=0.0$ (long-dashed line), $\sigma=1.25$ (dashed line), and $\sigma=2.0$ (solid line) are used in the calculations. (c) Velocity and porosity profiles obtained from a sediment core near Site-2.

uniform pore size assumption ($\sigma=0.0$) predicts a dispersion region below 10 kHz and there is no significant dispersion within the 10–80 kHz frequency band. The extended Biot theory predictions agree with the measurement results when silty-sand sediments, with pore-size standard deviations between $\sigma=1.25$ and $\sigma=2.0$, are considered. In Fig. 1(b), measured and calculated attenuation coefficients are compared. The measured attenuation values are in agreement with the theoretical predictions when sediments with distributed pore sizes are considered. Figure

Table 1. Physical properties and their values used in the calculation of sound speed and attenuation.

Physical Property	Symbol	Unit	Value
Grain density	ρ_r	kg/m^3	2650
Fluid density	ρ_f	kg/m^3	1024
Grain bulk modulus	K_r	N/m^2	3.6×10^{10}
Frame bulk modulus	K_s	N/m^2	3.69×10^{7}
Fluid bulk modulus	K_f	N/m^2	2.33×10^{9}
Shear modulus	μ	N/m^2	2.61×10^{7}
Dynamic viscosity of pore fluid	η	kg/m/s	1.0×10^{-3}
Permeability	k_s	m^2	5.0×10^{-11}
Porosity	β	⋯	0.46
Shear specific loss	δ_s	⋯	0.1
Frame volumetric specific loss	δ	⋯	0.05
Added mass coefficient	α	⋯	0.25

1(c) shows velocity and porosity profiles near Site-2, obtained from the sediment core analysis. At the probe depths (0.5 m), the sound velocities measured by a 500 kHz acoustic core logger are slightly higher than those measured by the *in situ* acoustic probe system. Grain size analysis at the probe depth (0.5 m) showed a mixture of gravel (11%), sand (70%), silt (12%), and clay (7%) at Site-2. Assuming a correlation between grain size and pore size distributions, this confirms our acoustic prediction of nonuniform ($\sigma \neq 0$) pore size distribution. The frequency dependency of the attenuation in sediments with various pore-size distributions will be elaborated in the next section.

3. Extended Biot theory predictions

Marine sediments can be regarded as a fluid-filled porous medium, and acoustic wave interaction with the bottom can be physically modeled by Biot's theory.[9] An extension of the Biot theory has been developed[8] for marine sediments with statistically distributed pore sizes and successfully validated for air-filled porous granular materials.[10] In the Biot theory, acoustic wave attenuation is primarily attributed to the viscous losses due to relative motion between the pore fluid and the skeletal frame. The viscous interaction is assumed to take place according to Darcy's law of fluid flow through porous media with the modification that the viscosity is made dependent on the frequency of the elastic wave. A frequency dependent viscosity correction factor, F, is defined by comparing the ratio of the total friction to the average fluid velocity in the oscillatory flow and steady laminar flow regimes. It is also assumed that the variation of viscous friction with frequency in a porous material follows the same laws as that in a single pore represented by a two-dimensional duct or a capillary tube of uniform cross section. Yamamoto and Turgut[8] have suggested a new viscous correction factor by calculating total viscous resistance and average seepage velocity for a macroscopic unit element with nonuniform pore size distribution. The new viscous correction factor is calculated as

$$F(\omega) = \frac{k_s}{\beta} \frac{2\int_0^\infty r^{-2} \kappa K(\kappa) e(r) dr}{\int_0^\infty [1 + 2iK(\kappa)/\kappa] e(r) dr}, \tag{4}$$

where $\kappa = r(\omega \rho_f / \eta)^{1/2}$ is a nondimensional parameter, ω is the angular frequency, k_s is the permeability, β is the porosity, r is the pore radius, $e(r)$ is the pore radius distribution function, η is the dynamic viscosity, ρ_f is the pore fluid density, and

$$K(x) = \frac{\mathrm{ber}'x + \mathrm{bei}'x}{\mathrm{ber}x + \mathrm{bei}\,x}, \tag{5}$$

in which ber and bei are the real and imaginary parts of Kelvin function and ber' and bei' are their derivatives.

In Fig. 2, pore size distributions of New Jersey Shelf sediments, measured by mercury injection technique, are shown. The pore-size distributions in Figs. 2(a) and 2(b) are approximated as log-normal [or ϕ-normal, $\phi = -\log_2 r$ (r in mm))] radius distributions with the standard deviation σ and mean $\bar{\phi}$ [$\sigma = 1.5$, $\bar{\phi} = 5.5$ in Fig. 2(a), and $\sigma = 2.2$, $\bar{\phi} = 7.8$ in Fig. 2(b)]. In Fig. 3, the measured and calculated phase velocity and attenuation curves are plotted in logarithmic scale to identify the velocity dispersion regions and to examine the frequency dependency of attenuation. In Figs. 3(a) and 3(b), it can be seen that, in sediments with nonuniform pore size distribution, the velocity dispersion and nonlinear attenuation regions due to viscous friction, are widened and shifted toward higher frequencies. Again, theoretical predictions by using $\sigma = 1.25$, and $\sigma = 2.0$ agree with the measured velocity and attenuation much better than that of using $\sigma = 0.0$. In Figs. 3(c) and 3(b), theoretical predictions of frequency dependency of attenuation are shown for $\sigma = 0.0$, $\sigma = 1.25$, and $\sigma = 2.0$. Using the definition $\alpha = kf^n$ (k is a constant and f is the frequency in kHz), the power exponents of $n = 0.6$, $n = 0.79$, and $n = 1.0$ are found by line fitting within the 10–100 kHz band. Similarly, within the 0.1–1 kHz frequency band, power exponent values of $n = 1.74$, $n = 1.45$, and $n = 1.35$ are found for $\sigma = 0.0$, $\sigma = 1.25$, and $\sigma = 2.0$, respectively. Sediments with sand, silt, and clay mixture are represented by higher values of σ,

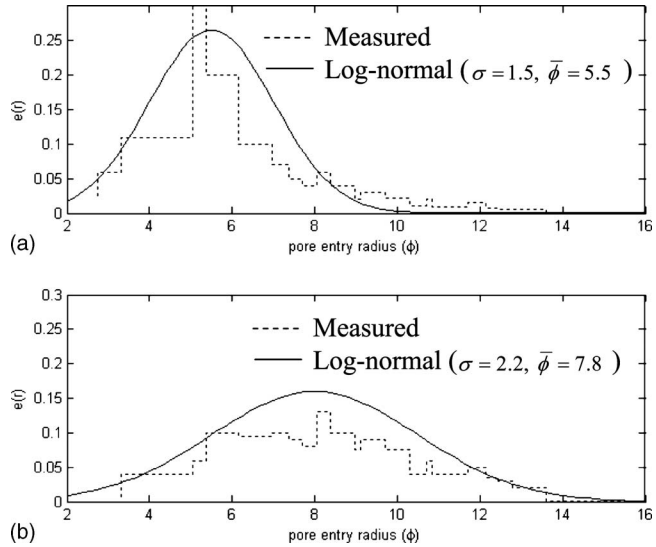


Fig. 2. Pore size distribution of New Jersey sediments measured by Mercury Injection technique (dashed lines). The log-normal pore-size distribution model with (a) σ=1.5, $\bar{\phi}$=5.5, and (b) σ=2.2, $\bar{\phi}$=7.8, are also shown (solid lines).

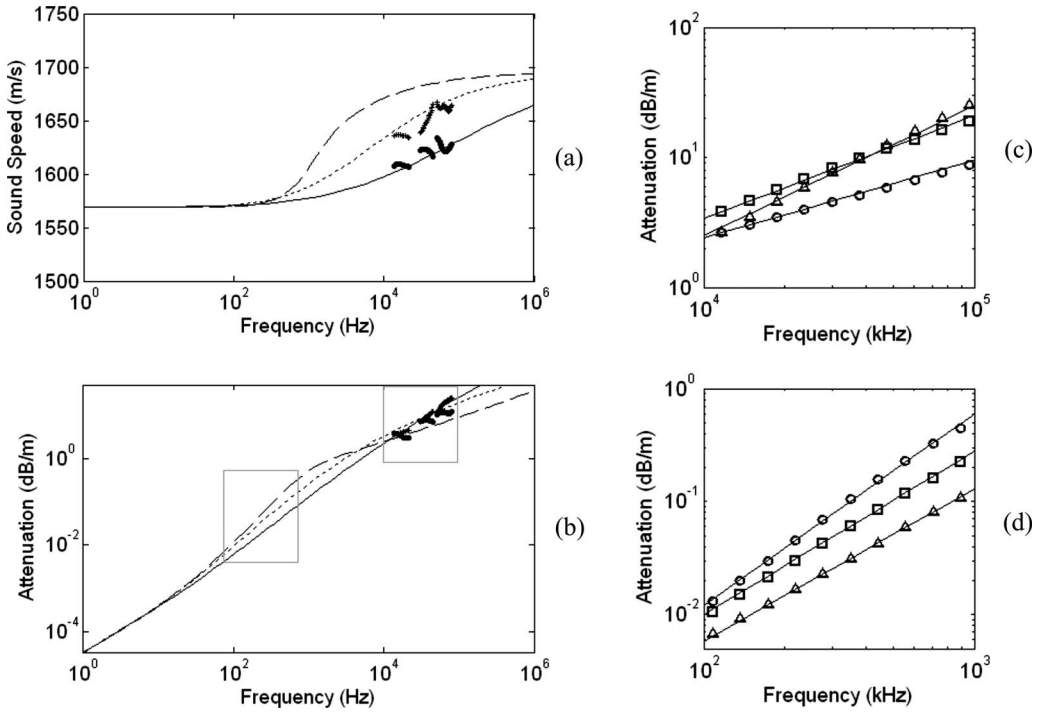


Fig. 3. (a-b) Comparison of measured velocity and attenuation curves with those calculated by using the extended Biot theory plotted in logarithmic scales [σ=0.0 (long-dashed line], σ=1.25 (dashed line), and σ=2.0 (solid line), (c-d) frequency dependency of attenuation at high- and low-frequency band, for σ=0.0 (circle), σ=1.25 (square), and σ=2.0 (triangle).

while sediments with more uniform grain sizes are represented by smaller value of σ, indicating more uniform pore sizes. Within the 0.1–1 kHz frequency band, the estimated power exponents are in good agreement with that of several low-frequency measurements, recently reported by Holmes *et al.*[11]

4. Summary and conclusions

In situ measurements of sound speed and attenuation were performed by using a wideband acoustic probe system, deployed at two silty-sand sites on the New Jersey Shelf. The spectral ratios of broadband pulses, transmitted between the 0.5-m-deep probes, showed phase velocity dispersion within the 10–80 kHz frequency band. The frequency region of observed velocity dispersion seems to be higher than that of previous measurements, performed in well-sorted beach sand and granular marine sediments.[1,3,5] As compared to the original Biot theory predictions, measured phase velocity dispersion results agree better with those predicted by the extended Biot theory with nonuniform pore size distribution. The observed linear frequency dependency of the attenuation also agrees with the theoretical predictions when sediments with nonuniform pore size distributions are considered.

Acknowledgments

This work was supported by the Office of the Naval Research. The authors thank the crew of the R/V Knorr for the excellent support during the 2006 Shallow Water Experiment. They also thank Chief Scientist Dr. D. J. Tang for the scientific support.

References and links

[1]A. Turgut and T. Yamamoto, "Measurements of acoustic wave velocities and attenuation in marine sediments," J. Acoust. Soc. Am. **87**, 2376–2382 (1990).
[2]A. Maguer, E. Bovio, and W. L. J. Fox, "Mechanisms for subcritical penetration into a sandy bottom: Experimental and modeling results," J. Acoust. Soc. Am. **107**, 1215–1225 (2000).
[3]R. D. Stoll, "Velocity dispersion in water-saturated granular sediment," J. Acoust. Soc. Am. **111**, 785–792 (2002).
[4]M. Buckingham and M. Richardson, "On tone-burst measurements of sound speed and attenuation in sandy marine sediments," IEEE J. Ocean. Eng. **27**, 413–428 (2002).
[5]J. C. Osler, A. P. Lyons, P. C. Hines, J. Scrutton, E. Pouliquen, D. Jones, D. M. F. Chapman, M. O'Connor, D. Caldwell, M. MacKenzie, I. B. Haya, and D. Nesbitt, "Measuring sound speed dispersion at mid to low frequency in sandy sediments: An overview of complementary experimental techniques developed for SAX04," In Underwater Acoustic Measurements: Technologies & Results, Heraklion, Crete, edited by John S. Papadakis and Leif Bjorno, 28th June—1st July (2005).
[6]C. W. Holland, J. Dettmer, and S. D. Dosso, "A technique for measuring in-situ compressional wave velocity dispersion in marine sediments," IEEE J. Ocean. Eng. **30**(4), 748–763 (2005).
[7]A. Turgut, R. Gauss, and J. Osler, "Measurements of velocity dispersion in marine sediments during the Baundary04 Malta plateau experiment," *Proceedings of the OCEANS 2005 MTS/IEEE, ISBN CD-ROM:* 0-933957-33-5 (2005).
[8]T. Yamamoto and A. Turgut, "Acoustic wave propagation through porous media with arbitrary pore size distributions," J. Acoust. Soc. Am. **83**, 1744–1751 (1988).
[9]M. A. Biot, "Theory of propagation of elastic waves in fluid-saturated porous solid: II High frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[10]K. V. Horoshenkov and M. J. Swift, "The acoustic properties of granular materials with pore size distribution close to log-normal," J. Acoust. Soc. Am. **100**(5), 2371–2378 (2001).
[11]J. D. Holmes, W. M. Carey, S. M. Dediu, and W. L. Siegmann, "Nonlinear frequency dependent attenuation in sandy sediments," J. Acoust. Soc. Am. **121**(5), EL218–222 (2007).

Observations of the R reflector and sediment interface reflection at the Shallow Water '06 Central Site

Jee Woong Choi
Department of Environmental Marine Sciences, Hanyang University, 1271 Sa-3-dong, Ansan, Korea
choijw@hanyang.ac.kr

Peter H. Dahl
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698
dahl@apl.washington.edu

John A. Goff
Institute for Geophysics, Jackson School of Geosciences, University of Texas at Austin, JJ Pickle Research Campus, Bldg. 196 (ROC), 10100 Burnet Rd. (R2200), Austin, Texas 78758-4445
goff@ig.utexas.edu

Abstract: Acoustic bottom-interacting measurements from the Shallow Water '06 experiment experiment (frequency range 1–20 kHz) are presented. These are co-located with coring and stratigraphic studies showing a thin (~20 cm) higher sound speed layer overlaying a thicker (~20 m) lower sound speed layer ending at a high-impedance reflector (R reflector). Reflections from the R reflector and analysis of the bottom reflection coefficient magnitude for the upper two sediment layers confirm both these features. Geoacoustic parameters are estimated, dispersion effects addressed, and forward modeling using the parabolic wave equation undertaken. The reflection coefficient measurements suggest a nonlinear attenuation law for the thin layer of sandy sediments.

PACS numbers: 43.30.Cq, 43.30.Ma, 43.30.Pc [WC]
Date Received: March 13, 2008 **Date Accepted:** May 31, 2008

1. Introduction

This paper presents results of measurements of bottom reflection made at frequencies 1–20 kHz, at location 39.0245 N, 73.0377 W (depth 80 m), near the shelf break on the New Jersey continental shelf. This location was the center point of a nominally 1 km^2 area defined as the central site for (mid-frequency) experimental observations [Fig. 1(a)] as part of the Shallow-Water '06 experiment, hereafter referred to as SW06.

Studies originating from previous experiments conducted on the New Jersey shelf, such as those involving the Shallow Water Acoustics in Random Media (SWARM) experiment site[1] and Atlantic Margin Coring Project (AMCOR) site,[2–4] offer potential comparisons with these results, in addition to providing the necessary background for understanding the marine geology and ocean acoustic properties of this continental shelf region. However, the SW06 results reported here involve a much higher frequency than those used in previous studies, and they are also highly localized to within a 0.3-km radius of the above location that is southwest of the vertical line array position of Woods Hole Oceanographic Institution (WHOI) during the SWARM experiment by 26 km, and southeast of the AMCOR borehole No. 6010 site by 7 km. Seabed heterogeneity on the New Jersey shelf is strong over these scales;[5] in particular, the AMCOR site is on a sand ridge, whereas the SW06 central site is on the clay-rich outer-shelf sediment wedge.[6] At this SW06 site, seafloor sand is confined to a thin (~20 cm) winnowed layer[5,7] rather than to a thicker, O(1) m, sand sheet typical of sand ridge sites.[5,6] For these reasons, the initial comparison of our results is limited to the direct *in situ* measurements of

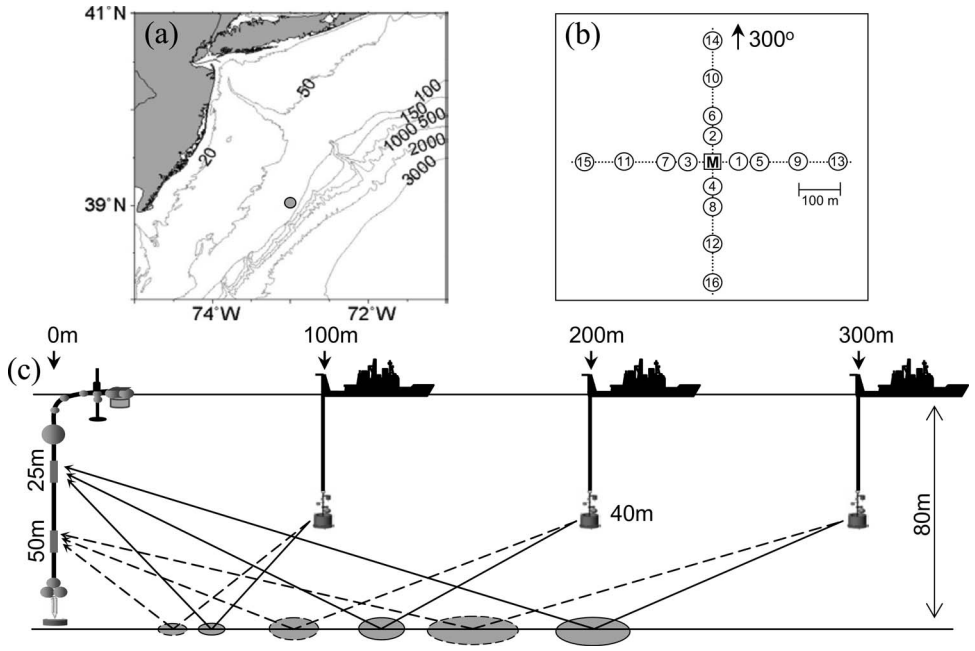


Fig. 1. (a) Experimental location on the New Jersey Shelf (circle) showing isobath contours in meters. (b) Experimental geometry showing 12 source stations along two transects. Source stations 1–4 are used in calibration. The moored receiving array is at the center of set of stations at location M. (c) Geometry showing source position (R/V KNORR) with respect to the receiving array and the resulting set of bottom grazing angles sampled. The ovals represent changing Fresnel zone size, e.g., at 10 kHz the Fresnel zone ranges between 4×4 m at range 100 m to 21×7 m at range 300 m for a receiver depth of 25 m.

sound speed made within the same 0.3 km radius[8] during SW06. Interpretation and forward modeling of our results are guided by the stratigraphic constraints provided by closely spaced (50 m) chirp seismic reflection profiles that provide pseudo three-dimensional coverage of the SW06 central site.[7]

Two kinds of acoustic observations are presented. The first represents a specific, and readily identifiable, single interaction observation of the R reflector, a regionally observed positive-impedance reflector.[7,9,10] Over the SW06 central site the R reflector is at a nominal depth of 22 m based on two-way travel time from the 1–4-kHz vertical incidence chirp data. The R reflector along our SW06 transect lines is relatively flat, changing by at most 3.5 ms (~3 m) over a span of 1000 m. The multipath corresponding to this reflector is seen in our data in the 1–4-kHz range, and by 6 kHz it has vanished into a background level owing to sediment attenuation. We verify these observations with simulation based on the parabolic equation algorithm.

The second observation is akin to bottom loss, or $-20\log_{10}|R|$, where R is the plane wave reflection coefficient for the seabed. These observations are from a single bottom bounce path that arrives before, and is time resolved from the signal associated with the R reflector. Analysis of arrival times using ray theory shows a perfect match in the timing of this bottom bounce path based on a waterborne path that is reflected once from the bottom. These measurements therefore represent a bottom loss measure restricted to surficial sediments *above* the R reflector. The surficial sediments at this site are fairly coarse (~1.0–1.3 ϕ medium-coarse sand), with high acoustic velocities (1720–1740 m/s) measured from *in situ* probes at 65 kHz.[5] Coring[7] reveals these coarse seafloor sediments to be confined within a thin (~20 cm) veneer covering a thicker (~20 m) layer of very clay rich and lower-velocity sediments (~1630–1660 m/s, measured at 257 kHz during core logging). The second observation is

therefore identified as a measurement of $-20\log_{10}|R_{13}|$, where R_{13} represents a partial or surficial layer reflection back to the water column (medium 1), from the thin layer (medium 2), and the intervening sediments (medium 3) between it and the R reflector.

2. Experimental description and observations

The acoustic observations were made by a group from the University of Washington Applied Physics Laboratory, from aboard the research vessel R/V KNORR (Knorr Leg-2 of SW06). An acoustic source was deployed at depth 40 m from the stern of the R/V KNORR, and signals were recorded on a moored receiving array system with remotely changeable receiving configuration; in this work signals received on the omni-directional receivers located at depths 25 and 50 m are analyzed. The receiving system was deployed at the above-mentioned coordinates, with this location henceforth referred to as location M.

Measurements were made at stations, defined as the stern position of the R/V KNORR while it underwent precise station keeping using its dynamic positioning system. The stations ranged between 100 and 300 m [Figs. 1(b) and 1(c)] from location M, resulting in a discrete set of six bottom grazing angles between 12.5° and 43.5°. The bearing angle between location M and the R/V KNORR for one transect of stations was 300°, and the other transects are offset this bearing angle by increments of 90°. Measurements were made over the course of the Fig. 1(b) geometry from 10–17 Aug. 2006 at all times of the day. Two types of pulses were used; one a 3-ms continuous wave (cw) pulse for which center frequencies between 4 and 20 kHz were superimposed and transmitted simultaneously on one source (spherical transducer) and the other a 5-ms cw pulse, for which center frequencies between 1 and 4 kHz were superimposed and transmitted simultaneously on another source (flooded-ring transducer, omnidirectional beam in horizontal, 80° beam width in vertical at 3 kHz) activated after a short (~1 s) delay. A particular frequency was recovered in postprocessing via digital bandpass filtering (acoustic data sampled at 50 kHz sampling rate).

For the $-20\log_{10}|R_{13}|$ estimates, the measurements are interpreted as the total acoustic field associated with a *single interaction* with the seabed. This measure represents an average of the squared envelope of received voltage (pressure) taken over 20 pings and averaging in this manner provides an estimate of the squared magnitude of the flat-interface reflection coefficient.[11] An *in situ*, through-the-system, calibration was carried out at four stations located at range 50 m, each separated by 90° in bearing angle (stations 1–4 in Fig. 1). The calibration yielded an estimate of a single, integrated system parameter (ISP) and the variance of ISP (over the course of five days of ISP measurements and over the four bearing angles) forms the major component of measurement uncertainty for estimates of $-20\log_{10}|R_{13}|$. The conductivity-temperature-depth measurements from the R/V KNORR were used to compute ray-based estimates of transmission loss (TL) and seabed grazing angle, θ_g. As a check on the stability of the eigenray paths due to changing sound speed profile, a ray analysis was done for each day from 10–15 Aug. using 15-min-averaged sound speed profiles derived from temperature measurements from a nearby WHOI mooring, generated every 30 s. The analysis showed TL to vary by <0.5 dB and θ_g to vary by $<0.5°$, for the bottom bounce eigenray paths, confirming that our measurements "slip under" and were otherwise not influenced by time-varying water column properties.

Figure 2 summarizes the acoustic observations and displays one of two main results of this paper. A typical sound speed profile [Fig. 2(a)] and the corresponding ray diagram [Fig. 2(b)] show the first six eigenrays delivering the signal over the 200-m range transmission via waterborne paths. These paths in their order of arrival (for 25 m depth receiver) are the direct (D), surface (S), bottom (B), bottom-surface (BS), surface-bottom (SB), and surface-bottom-surface (SBS). The bottom grazing angle, θ_g, for the B path is 25° (19.5° for the 50 m depth receiver). An additional sediment-borne path associated with the R reflector, or R path (R), is depicted in the illustration [Fig. 2(c)].

A time series of the relative received level for a pulse with center frequency 2 kHz [Fig. 2(d)] shows arrival time structure associated with the above paths displayed in Figs. 2(b) and 2(c). The 2-kHz results show a strong R path and additional faint arrivals that we postulate

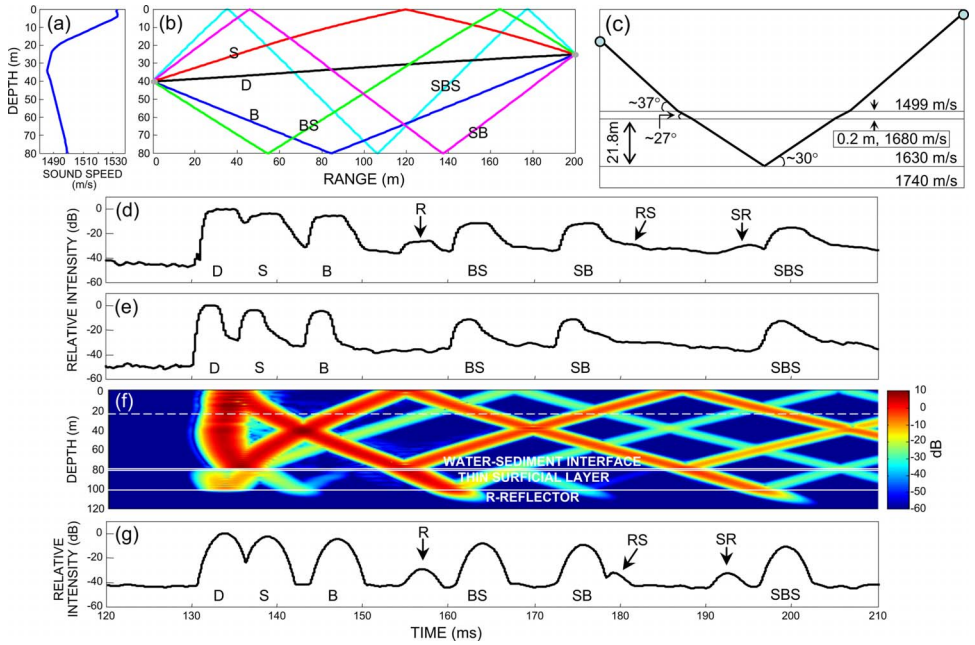


Fig. 2. (a) (Color online) Representative sound speed profile for 10 Aug. 11:07 UTC. (b) Corresponding ray diagram for a source at 40 m, 25 m receiver depth, and range 200 m, showing the first six eigenrays. The third arriving eigenray is the bottom bounce path (B) for which an estimate of $-20\log_{10}|R_{13}|$ is made. Other waterborne paths are the direct (D), surface (S), bottom-surface (BS), surface-bottom (SB), and surface-bottom surface (SBS). (c) An illustration of the sediment-borne path associated with the R reflector (R); the angles noted apply to case of source at 40 m, 25-m receiver depth and range 200 m. (d) Time series of received level for 2-kHz center frequency, based on the average of 20 ping transmissions made on 10 Aug at 10:00 UTC, with acoustic source at station 10 as shown in Fig. 1(b). (e) Simultaneously measured time series of received level for 6 kHz center frequency; here the R path has vanished into the intensity level formed by time spreading of other paths and sediment attenuation. (f) PE-simulated acoustic field (center frequency 2 kHz) for this geometry (see text for description of geoacoustic model used). (g) PE-simulated time series for a source depth at 40 m, receiver depth 25 m, and range 200 m, showing the R-reflector multi-path (R) and additional multi-path species of R-reflector-surface (RS) and surface-R-reflector (SR). Noise has been added to the time series to mimic the nominal, expected ratio for signal-to-background level.

to be R surface (RS) and surface R (SR) paths. In contrast to the 2-kHz results, the R path is not observable in the simultaneously measured 6-kHz results [Fig. 2(e)] as sediment attenuation places this arrival beneath the background level formed by time spreading of prior-arriving paths plus additive noise. The received levels are arbitrarily set to 0 dB for the D path, and an estimate of $-20\log_{10}|R_{13}|$ for the B path is 0.7 dB for 2 kHz and 1.4 dB for 6 kHz.

A calculation performed with a parabolic wave equation (PE) code[12] for this geometry [Fig. 2(f)] shows the acoustic field versus depth and time for a source at depth 40 m, using a 5-ms pulse with center frequency 2 kHz. For the simulation the water column sound speed profile [Fig. 2(a)] is used together with the geoacoustic model in Fig. 2(c). Note: the thin layer sound speed of 1680 m/s corresponds to a dispersion-corrected compressional wave speed at 2 kHz (obtained from Fig. 3 in Ref. 13) of the 1730 m/s speed as determined from *in situ* acoustic measurements at 65 kHz (i.e., mean value of 1720–1740 m/s as mentioned in the introduction). However, in view of the 20-cm-thick layer and 2 kHz frequency, both speeds produce nearly identical results. The most important parameters are the large-layer depth (21.8 m) and speed within this layer (1630 m/s). These are determined from our analysis of the travel time difference between the B and R paths [e.g., as shown in Fig. 2(d)] measured at ranges 200 and 300 m, with an uncertainty in layer depth of ±1 m and speed of ±20 m/s. The densities for the surficial and second layers are assumed to be 2.1 g/cm^3 and 2 g/cm^3 based on core logs made in this area.[14] An empirical relation for compressional wave attenuation within the surfi-

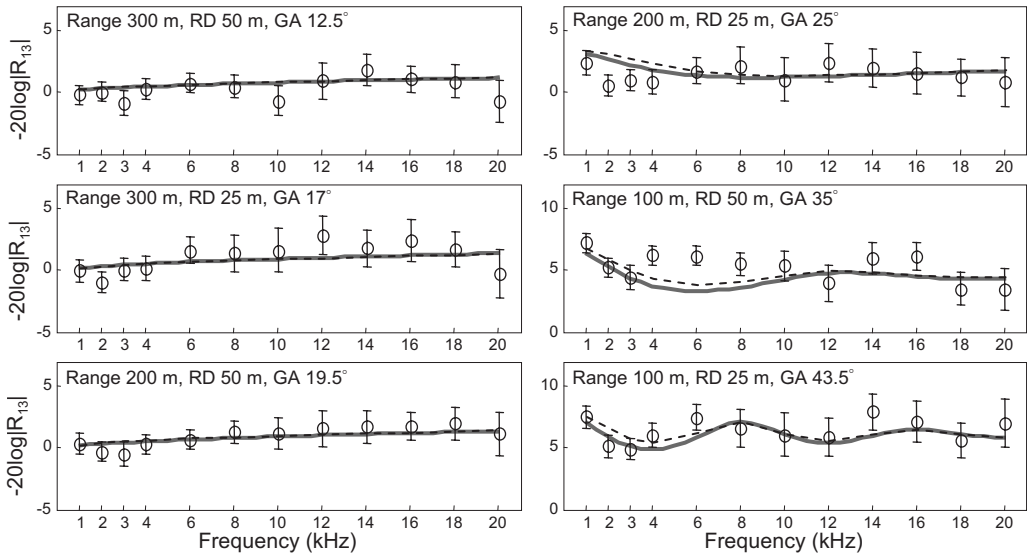


Fig. 3. Measurements of $-20\log_{10}|R_{13}|$ as a function of frequency for six grazing angles (GA) between 12.5° and 43.5°. The grazing angle, source-receiver ranges, and receiver depths (RD) associated with each grazing angle are noted at the top of each plot. The measurements are compared to a two-layered fluid sediment model for which surficial sediment sound speed in the upper (20 cm) layer is 1730 m/s (gray line) and depends on frequency (dashed line) according to a dispersion correction applicable to coarse sand. The frequencies 1, 2, 3, 4, 6, 8, 10, 12, 14, 16, 18, and 20 kHz use 1650, 1680, 1695, 1704, 1711, 1716, 1720, 1723, 1725, 1726, 1727, and 1728 m/s, respectively. Other geoacoustic parameters are discussed in the text.

cial sediment is taken to be $0.2(f/f_{\text{ref}})^{1.6}$ dB/m, where f is a frequency in kHz ($f_{\text{ref}}=1$ kHz), which is a result of the analysis of measurements of $-20\log_{10}|R_{13}|$ as discussed further below and is limited to the frequency range 1–20 kHz. This relation falls within the nominal envelope of attenuation data from sandy sediments corresponding to this frequency range.[11,15]

The attenuation within the second layer is taken to be 0.05 ± 0.01 dB/m/kHz, a value estimated by examining the ratio of amplitudes between the B and R paths for frequencies 1, 2, 3, and 4 kHz, accounting for differences in waterborne TL, assuming reflection and transmission from the three interfaces [shown in Fig. 2(c)] is constant within this narrow frequency band, and taking the total path length within the second layer [Fig. 2(c)] to be 87 m. The attenuation for the thin (20 cm) layer is significantly greater than that used for the larger (~20 m) layer, however a larger attenuation is expected in view of the coarse sand composition of the thin layer.[15]

Finally, the R reflector itself is modeled as a half space with a compressional wave speed of 1740 m/s, density of 2.2 g/cm^3, and attenuation of 0.3 dB/m/kHz, which are taken from inverted values from SW06.[16] The PE calculation clearly shows a reflected field emerging from the layer depth at 22 m—or the R reflector. A cut from this [dashed line in Fig. 2(f)] provides a simulated time series for a receiver depth of 25 m [Fig. 2(g)] that is comparable with data [Fig. 2(d)]. Interestingly, multiple species of the R path, e.g., R-surface (RS), and surface-R (SR) can be seen in both simulation and (faintly) in the 2-kHz data [Fig. 2(d)]. The 2-kHz data and PE simulated time series compare well in terms of timing and arrival structure, supporting a geoacoustic description of the seabed consisting of a thin (~20 cm) surficial layer of higher compressional sound speed, over thicker (~22 m) layer of sediment with slightly lower compressional speed that lies above a higher speed (impedance) reflector, as shown in Fig. 2(c).

The second main result (Fig. 3) are the estimates of the $-20\log_{10}|R_{13}|$ as a function of frequency for the six grazing angles available from the geometry shown in Fig. 1(c) and simul-

taneously measured frequencies between 1 and 20 kHz. The aforementioned two-layer model is compared with these data in two ways. The first (gray line) utilizes 1730 m/s in the thin layer, and second (dashed line) applies a frequency-dependent dispersion correction[13] applicable to coarse sand, for which the sound speed in the thin layer ranges from 1650 m/s at 1 kHz to 1728 m/s at 20 kHz (see figure caption). Both ways utilize the arrival-time inverted estimate of 1630 m/s for the sediments below the thin layer, and the density profiles mentioned in context of Fig. 2. The sediment region below the thin layer is treated as a half space, as the R reflector is time resolved from and not adding to the bottom bounce path, and any impedance change at equivalent subseafloor depths of ~20 m cannot be seen in modeling results at these frequencies. Measurements at the three shallow grazing angles are most sensitive to attenuation within the thin surficial layer, and a nonlinear property of attenuation is suggested by the minor upward slope in $-20\log_{10}|R_{13}|$ estimate with increasing frequency. It is found that the above-mentioned $0.2(f/f_{\mathrm{ref}})^{1.6}$ dB/m relation for the surficial sediment attenuation in this layer provides the best fit to the data.

The dispersion correction yields modest, if any, improvement in view of the variance of the measurements. However, the estimates of $-20\log_{10}|R_{13}|$ are very consistent with the presence of a thin, 20-cm layer overlying a half-space speed of 1630 m/s. Two direct (ground truth) measurements of sound speed were made within a 50 m radius of station 12 [see Fig. 1(b)],[8] using a 2–11-kHz low-frequency (LF) and 10–21-kHz mid-frequency (MF) probe pulse in each case averaging to a depth of 1.6 m into seafloor. In one case the LF and MF speeds were estimated as 1615 and 1622 m/s, respectively, and in the other the LF and MF speeds were estimated as 1598 and 1599 m/s, respectively, with an uncertainty of approximately ±10 m/s applying to all estimates. Given that the 20-cm layer constitutes about 12% of this instrument's averaging depth it is reasonable to assume that these sound speed estimates apply to the region below the 20-cm layer, and are consistent with our corresponding estimate of 1630±20 m/s.

3. Conclusions

The acoustic bottom-interacting measurements from SW06 reported here provide a clear demonstration of the role of stratigraphic constraints and ground truth data on sediment bulk physical properties, on both geoacoustic inversion and acoustic forward modeling. The acoustic measurements made between 1 and 20 kHz are highly localized (within a radius of 300 m) and co-located coring and stratigraphic studies show a thin (~20 cm) higher sound speed layer overlaying a thicker (~20 m) lower sound speed layer ending at a high-impedance reflector (R reflector). The acoustic measurements yielded two key observables: (1) direct measurements of the reflections from the R reflector (for <6 kHz) and (2) estimates of $-20\log_{10}|R_{13}|$ (for 1–20 kHz) from a single bottom bounce path that arrives before, and is time resolved from, the signal associated with the R reflector. In terms of inversion, the R reflector travel time analysis yielded an estimate of the thick layer depth to be 22±1 m within which the compressional wave speed and attenuation were 1630±20 m/s and 0.05±0.01 dB/m/kHz, respectively. Forward modeling using the parabolic equation algorithm reproduced well the arrival structure at 2 kHz.

In contrast, the estimates of $-20\log_{10}|R_{13}|$ are more sensitive to the aforementioned thin, higher speed layer, and the data suggest a nonlinear attenuation law in sandy sediment[17] is more appropriate than a linear one as indicated by the minor upward slope in $-20\log_{10}|R_{13}|$ with increasing frequency at low grazing angles. For the underlying clay-rich sediment layer, a linear frequency-attenuation was estimated by examining the ratio of amplitude between the B and R paths for the 1–4-kHz frequency band, and was utilized in the modeling of $-20\log_{10}|R_{13}|$ for 1–20 kHz within this layer. We do not insist that a linear-frequency dependence apply to the entire 1–20-kHz band, although physical reasons support a linear assumption for such sediments.[17] Finally, and of considerable importance in terms of consistency, the inversion result from the R reflector reflection data, and the modeling result for the $-20\log_{10}|R_{13}|$ estimates, were both reasonably consistent with the co-located direct measurements of sediment sound speed to a depth of 1.6 m.

Acknowledgments

This research was supported by the Office of Naval Research. The mooring No. 54 data used for simulation was made available courtesy of the Woods Hole Oceanographic Institution, Ocean Acoustics Laboratory. UTIG contribution 1989.

References and links

[1]Y.-M. Jiang, N. R. Chapman, and M. Badiey, "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data," J. Acoust. Soc. Am. **121**, 1879–1894 (2007).
[2]M. V. Trevorrow and T. Yamamoto, "Summary of marine sedimentary shear modulus and acoustic speed profile results using a gravity wave inversion technique," J. Acoust. Soc. Am. **90**, 441–456 (1991).
[3]W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey continental shelf," IEEE J. Ocean. Eng. **20**, 321–336 (1995).
[4]D. P. Knobles, E. K. Westwood, and J. E. LeMond, "Modal time-series structure in a shallow-water environment," IEEE J. Ocean. Eng. **23**, 188–202 (1998).
[5]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[6]T. A. Davies, J. A. Austin, Jr., M. B. Lagoe, and J. D. Milliman, "Late Quaternary sedimentation off New Jersey: New results using 3-D seismic profiles and cores," Mar. Geol. **108**, 323–343 (1992).
[7]C. Fulthorpe and J. A. Austin, "Shallowly buried, enigmatic seismic stratigraphy on the New Jersey outer shelf: Latest Pleistocene catastrophic erosion?," Geology **32**, 1013–1016 (2004).
[8]J. Yang, D. J. Tang, and K. L. Williams, "Direct measurement of sediment sound speed in Shallow Water '06," J. Acoust. Soc. Am. **124**, EL116–EL121 (2008).
[9]S. P. S. Gulick, J. A. Goff, J. A. Austin, Jr., C. R. Alexander, Jr., S. Nordfjord, and C. S. Fulthorpe, "Basal inflection-controlled shelf-edge wedges off New Jersey track sea-level fall," Geology **33**, 429–432 (2005).
[10]J. D. Milliman, A. Jiezao, L. Anchun, and J. I. Ewing, "Late Quaternary sedimentation on the outer and middle New Jersey continental shelf: Result of two local deglaciations," J. Geol. **98**, 966–976 (1990).
[11]D. R. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics* (Springer, New York, 2006).
[12]M. D. Collins, "A split-step Padé solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[13]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).
[14]J. A. Goff, J. A. Austin, Jr., B. Christensen, and A. Turgut, "Chirp seismic reflection data on the New Jersey middle and outer shelf: The geologic response to 40,000 years of sea level change (A)," J. Acoust. Soc. Am. **122**, 2983 (2007).
[15]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1340 (1980).
[16]Y.-M. Jiang and N. R. Chapman, "Bayesian geoacoustic inversion in a dynamic shallow water environment," J. Acoust. Soc. Am. **123**, EL155 (2008).
[17]J. D. Holmes, W. M. Carey, S. M. Dediu, and W. L. Siegmann, "Nonlinear frequency-dependent attenuation in sandy sediments," J. Acoust. Soc. Am. **121** EL218–EL222 (2007).

Short range travel time geoacoustic inversion with vertical line array

Yong-Min Jiang and N. Ross Chapman
School of Earth and Ocean Sciences, University of Victoria, PO Box 3055 Victoria, British Columbia, Canada, V8W 3P6
minj@uvic.ca, chapman@uvic.ca

Peter Gerstoft
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238
gerstoft@ucsd.edu

Abstract: This paper presents travel time geoacoustic inversion of broadband data collected on a vertical line array at short range of 230 m during the Shallow Water 2006 experiments. A ray-tracing method combined with a hybrid optimization algorithm that utilizes differential evolution and downhill simplex was used for the inversion of sediment properties. The ocean sound speed profile and geometric parameters were inverted prior to the sea bottom properties to account for the temporally variable ocean environment. The sediment sound speed and thickness estimates are consistent with *in situ* measurements and matched-field inversion results of longer-range data from the experiment.

PACS numbers: 43.30.Pc, 43.60.Pt [WC]
Date Received: March 17, 2008 **Date Accepted:** June 9, 2008

1. Introduction

Many geoacoustic inversion techniques have been developed and successfully implemented to estimate sea bottom properties. Most of the research assumed that the ocean sound speed profile (SSP) was spatio-temporally invariant. However, the ocean SSP has significant impact on matched-field geoacoustic inversion (MFI),[1] and MFI performance is improved by including ocean sound speed variability in the inversion.[2,3]

The objective of this work is to report travel time geoacoustic inversion in a time varying ocean environment. The data were collected on the Marine Physical Laboratory vertical line array (MPL-VLA1) during the Shallow Water 2006 (SW06) experiments. SW06 was a multidisciplinary, multi-institution experiment that was carried out near the shelf break on the New Jersey continental shelf from mid-July to mid-September, 2006.[4] The objectives of the geoacoustic inversion component of SW06 were to estimate seabed properties at a common interest site through different approaches (i.e., using different acoustic field information and inversion algorithms), compare the limitations of each method, and understand the influence of the oceanographic variability on the estimated parameters and their uncertainties.

The oceanographic variation was observed from the SSPs measured at the source position. In order to reduce sea bottom estimation uncertainties arising from temporal variation of the ocean environment, the ocean SSP was parameterized in terms of empirical orthogonal functions (EOFs) and their coefficients were inverted prior to the geoacoustic inversion. The acoustic model used here is based on ray theory. The inversion combined the global optimization algorithm differential evolution[5] and the local optimization algorithm downhill simplex[6] (adaptive simplex differential evolution, ASDE) to estimate parameters of a simple geoacoustic model. The experimental design with multiple source depths and multiple hydrophones at close range provided the multiplicity of sub-bottom acoustic paths to resolve the sediment properties.

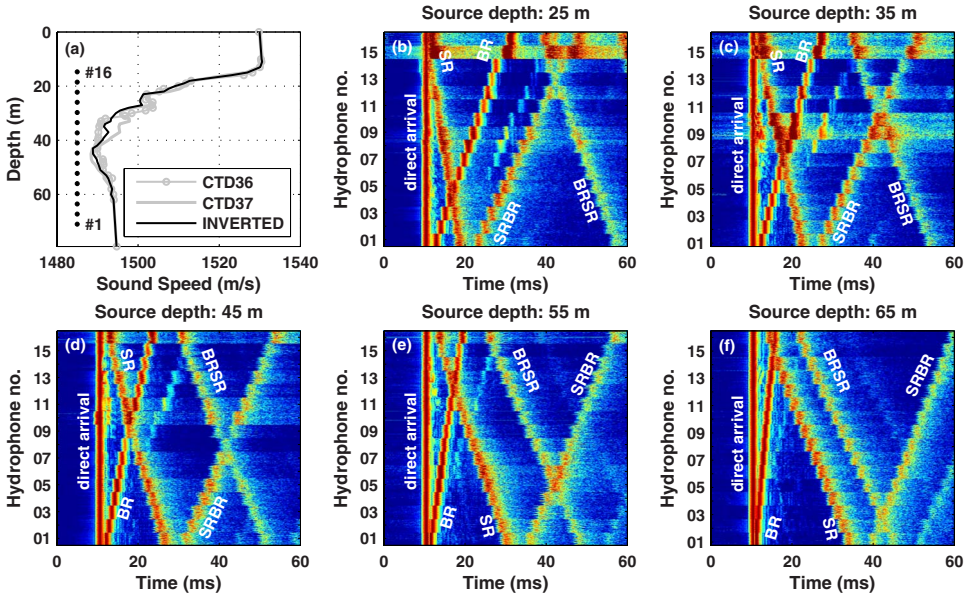


Fig. 1. (Color online) (a) CTD36, CTD37 and inverted SSP for the 25 m source data. The black dots indicate the hydrophone positions in the water. (b)-(f) Normalized matched filtered LFM signals collected on 16 hydrophones for the 25–65 m source depths.

2. Short range acoustic experiment

The data were collected on 31 August on MPL-VLA1 that was deployed at (39°1.4771′N, 73°2.259′W). Sixteen hydrophones were equally spaced on the array with a separation of 3.75 m, and the bottommost hydrophone (No.1) was 8.2 m above the sea floor. The source was deployed from the R/V Knorr, lowered down in the water column from 15 to 65 m in 10 m intervals, and was held for 5 min at each depth. During the signal transmission, the ship was maintained at ~230 m from the VLA using dynamic positioning. The transmitted signal was a 1 s 1500–4500 Hz mid-frequency linear frequency modulation pulse, repeated each second.

The bathymetry between the source and MPL-VLA1 is almost flat with 79.6 m at the source and 79.0 m at MPL-VLA1. CTDs (conductivity-temperature density) were measured immediately before (CTD36) and after (CTD37) the experiment. The time between the two CTDs was 45 min. The derived SSPs shown in Fig. 1(a) have almost iso-velocity layers at the top and lower portions of the water column, and a strong negative downward refraction thermocline from 15–30 m. The middle part of the SSP shows strong time variation.

Mid-frequency LFM signals received at all of the hydrophones on the VLA were matched filtered by the replica source pulse. To eliminate undesired noise in the data, the signal bandwidth was limited to 1500–3800 Hz. The envelopes of the normalized matched filtered signals (dB) are shown in Figs. 1(b) to 1(f). Each panel represents 1 min signals (60 transmissions stacked vertically) received at all hydrophones at the source depths from 25 to 65 m. The 60 transmissions shown for each hydrophone give a qualitative impression of the signal variability.

Ray tracing[7] was used to identify the direct, bottom reflected (BR), surface reflected (SR), surface reflected-bottom reflected (SRBR), and bottom reflected-surface reflected (BRSR) arrivals. For all the source depths, the BR paths at the hydrophones in the bottom portion of water column have small travel time variations. SR paths at the hydrophones in the top portion of the water column have greater amplitude and travel time variations due to source

motion, sea surface roughness, and relatively larger hydrophone position variation. Sub-bottom reflections are observed between BR and SR/SRBR paths for the 45, 55, and 65 m sources, or between BR and SRBR for the 25 and 35 m sources.

Multiple direct arrivals were observed for the sources and receivers in the thermocline (15–30 m) due to the strong negative gradient and the existence of microstructures in the thermocline layer. The direct path amplitude from the ray tracing was not accurate due to the multipath interference, so there was no attempt to extract bottom reflection coefficients. Instead we used the travel times for inversion.

3. Acoustic model, objective function, and bottom parameterization

A range independent ray-tracing model to invert for the sub-bottom geoacoustic profile was developed for this study. The model searched for the ray parameter $[\cos(\varphi(z))/c(z)$, where φ, c, and z are grazing angle, sound speed and depth] for an eigenray that connected the source and receiver within any given range tolerance (here 0.005 m). The inversion was done in two stages, the first to invert for the SSP and geometry (water depth, range, source and receiver depths, array tilt) of the experiment, and then the geoacoustic parameters.

The inversion algorithm is a hybrid optimization method, adaptive simplex differential evolution ASDE. The local optimization downhill simplex algorithm[6] is embedded in the global optimization algorithm differential evolution.[5] The perturbation size in the downhill simplex is adaptively adjusted according to the models being accepted. Control parameters for ASDE are: population size[5] is 20 times of the number of parameters to be inverted; mutation factor is 0.8; crossover factor is 0.8; and the perturbation number in the downhill process is 5.

The objective function compares the differences of the modeled ray travel time $T_{i,j}(\bar{m})$ and the measured ray travel time $t_{i,j}$: $E(\bar{m})=\Sigma_{i=1}^{N_L}\Sigma_{j=1}^{N_H}(t_{i,j}-T_{i,j}(\bar{m}))^2$, where $\bar{m}$ is the vector of model parameters to be inverted; N_H is the number of hydrophones; and N_L is the number of reflectors used in the inversion. When tracing rays reflected from ocean surface and bottom, $N_L=2$ while $t_{i,j}$ represents the time differences between BR paths and direct arrivals, and the time differences between SR paths and direct arrivals. When tracing rays reflected from the sub-bottom layers, N_L equals the number of layers and $t_{i,j}$ represents the time difference between the sub-bottom reflection path and the corresponding BR path.

The ability to resolve bottom structures is limited by the time resolution of the transmitted signal and the strength of the sub-bottom reflections. The bottom parameterization was based on resolvable sub-bottom reflections. Figure 1 shows a strong bottom reflection for all five source depths (no sub-bottom reflection was identified for the 15 m source). For the 55 and 65 m source depths, multiple sub-bottom layer reflections are also revealed by a weak signal enhancement technique (similar to automatic gain control) that was applied to the matched filtered time series. Figure 2(b) shows the enhanced 55 m source depth signals. The multiple hydrophone geometry provided a multiplicity of acoustic paths in the sub-bottom layer for the travel time inversion.

4. Effect of water column SSP variations to the received signal

Figure 2(a) shows the envelope of the matched filtered time series for the 25 m source depth. The resolution of the signal is about 0.5 ms. Figure 2(a) compares the travel time predictions using the previously known values for the experimental geometry and the two ocean SSP measurements, and the inverted experimental geometry and SSP. The travel time difference between predictions using CTD36 or CTD37 for hydrophones 8–14 are up to 2 ms. This is more than the resolution of the signal, thus the measured SSPs are not accurate enough for the geoacoustic inversion.

The direct, sea surface, and bottom reflection paths in the water column were used to invert for the experimental geometry and ocean SSP. Two steps were taken to improve the estimates of these parameters for the geoacoustic inversion. First, the 65 m source signals were used to invert for the experimental geometry using CTD37, because this CTD was obtained immediately after the 65 m source transmission. The inverted values were 79.2 m for water

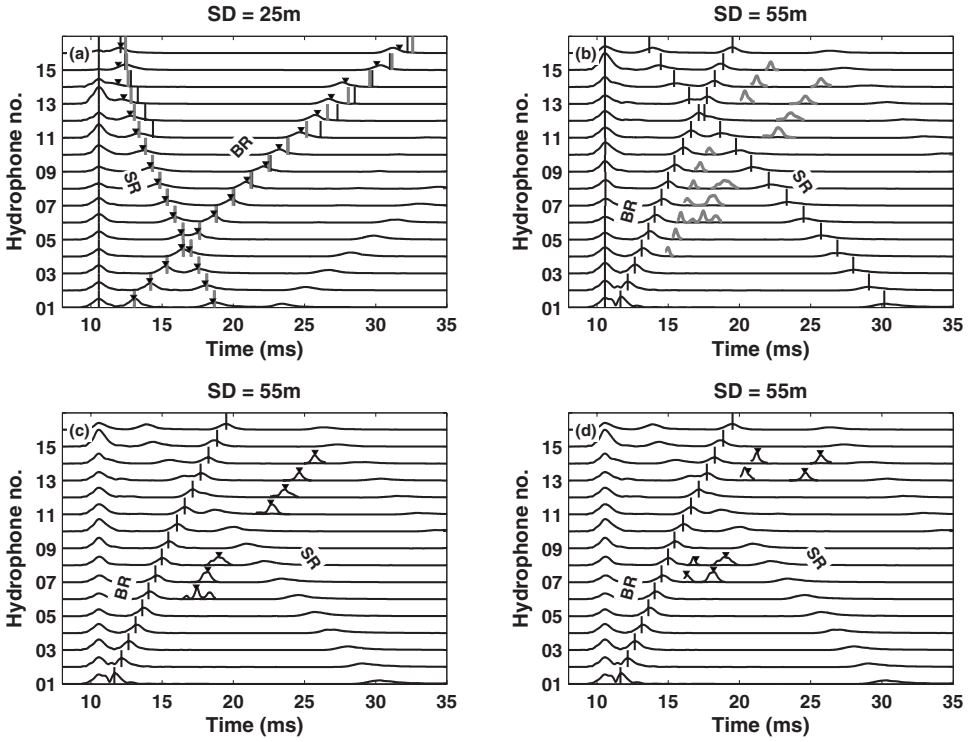


Fig. 2. Received waveforms for the (a) 25 m and (b)–(d) 55 m source with enhanced sub-bottom reflections for the 55 m source data [gray curves in (b)]. The travel times superimposed in (a) are obtained using CTD36 (vertical gray lines), CTD37 (vertical black lines), and inverted SSP (black triangles); and in (b) using the inverted ocean SSP (vertical lines). The travel times produced by inverted geoacoustic parameters for (c) one-layer model and (d) two-layer model are marked in black triangles.

depth, 233.4 m for range, 64.6 m for source depth. The search bounds were $[78,81]$ m, $[220,240]$ m, and $[60,70]$ m, respectively. Second, SSPs for the other source depths were parameterized by EOFs,[3] and the EOF coefficients were inverted using the established geometric parameters. The inverted SSP for 25 m source data is shown in Fig. 1(a), and the travel time predictions obtained using the inverted geometry and SSP for the 55 m source data are also shown in Fig. 2(b).

5. Inversion of sediment sound speed and layer thickness

The sub-bottom reflection in the 25–65 m source data which is apparent in Fig. 1 suggests the existence of a strong reflector in the bottom. The inversion first assumes a single sediment layer over a half space. The inversion results and the corresponding two way travel times (TWTs) are shown in Table 1. For the 55 and 65 m source data, a two-layer sediment over a half space was

Table 1. Travel time inversion results for one-layer over half space model.

Parameter	Source depth 25 m	35 m	45 m	55 m	65 m
Layer thickness (m)	21.5	20.6	22.3	22.3	20.4
Sound speed (m/s)	1609	1597	1600	16001	1584
TWT (ms)	26.7	25.8	27.9	27.9	25.8

Table 2. Travel time inversion results for two-layer over half space model.

	Layer I		Layer II	
Source depth	Sound speed	Layer thickness	Sound speed	Layer thickness
55 m	1581 m/s	11.9 m	1602 m/s	9.1 m
65 m	1584 m/s	14.5 m	1611 m/s	8.4 m

also used in the inversion. The results are shown in Table 2. In all the inversions, the search bounds were 1550–1700 m/s for sound speed and 5–30 m for the layer thickness

Sediment sound speed and layer thickness estimates of a one-layer model for all five source depths are consistent with each other (see Table 1). For the two-layer model, Table 2, the sound speed estimates are consistent with each other, and the average sound speed is consistent with the one-layer inversion results. The estimated values provide accurate predictions of the travel times of the sub-bottom reflections [Figs. 2(c) and 2(d)]. The estimates are also consistent with the results of matched-field inversion of longer range data along the same radial path from MPL-VLA1.[2,3] The inverted sound speeds represent the average sound speeds in the layers.

The SW06 experiment site has been extensively surveyed previously so that a large amount of geological and geophysical ground truth information and oceanographic observations are available. Many geoacoustic inversion studies have been done on the New Jersey Shelf previously. Due to the variability of the seabed properties from location to location,[8] we only compare our estimates with most recent relevant works[8–10] near MPL-VLA1. Overall, the sound speed estimates are consistent with *in situ* sound speed measurements[9] made about 250 m from MPL-VLA1 (1615±9 m/s and 1598±11 m/s) in the frequency band 2–11 kHz. The combination of sediment sound speed and layer thickness is also consistent with chirp seismic reflection data in terms of two way travel time (26.1±0.5 ms) from the sea bottom down to the "R" reflector.[8]

6. Conclusion

Travel time geoacoustic inversion was applied to VLA data collected at a source range of 230 m for a site on the New Jersey continental shelf. The short-range, varying-source-depth and VLA geometry provided wide grazing angle coverage of bottom interaction, as well as a multiplicity of sub-bottom propagation paths through the layers for the geoacoustic inversion. The effect of time varying ocean sound speed was mitigated by inverting for the ocean SSPs prior to the geoacoustic inversion. The hybrid optimization algorithm ASDE is efficient for estimating the model parameters.

Acknowledgments

This work is supported by the Office of Naval Research under Grant Nos. N00014-05-1-0264 and N00014-03-1-0131. The authors would like to thank Dr. William S. Hodgkiss, for the acoustic data, Dr. David Knobles for the navigation data, and Dr. John Goff for the chirp seismic reflection data.

References and links

[1]M. Siderius, P. L. Nielsen, J. Sellschopp, M. Snellen, and D. Simons, "Experimental study of geo-acoustic inversion uncertainty due to ocean sound-speed fluctuations," J. Acoust. Soc. Am. **110**, 769–781 (2001).
[2]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Effect of ocean sound speed uncertainty on matched-field geoacoustic inversion," J. Acoust. Soc. Am. **124**, EL162–EL168 (2008).
[3]Y.-M. Jiang and N. R. Chapman, "Bayesian geoacoustic inversion in a dynamic shallow water environment," J. Acoust. Soc. Am. **123**, EL155–EL161 (2008).
[4]D. J. Tang, J. Moum, J. Lynch, P. Abbot, R. Chapman, P. Dahl, T. Duda, G. Gawarkiewicz, S. Glenn, J. Goff, H. Graber, J. Kemp, A. Maffei, J. Nash, and A. Newhall, "Shallow Water '06-A joint acoustic propagation/nonlinear internal wave physics experiment," Oceanogr. **20**, 156–167 (2007).
[5]R. Storn and K. Price, "Differential evolution–a simple and efficient adaptive scheme for global optimization over continuous spaces," Technical Report No. TR-95-012, ICSI, March 1995.
[6]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran*, 2nd ed.

(Cambridge University Press, Cambridge, 1992), pp. 402–406.
[7]M. B. Porter and Y.-C. Liu, "Finite element ray tracing," in *Theoretical and Computational Acoustics*, edited by D. Lee and M. H. Schultz (World Scientific, Singapore, 1994), Vol. **2**, pp. 947–953.
[8]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[9]J. Yang, D. J. Tang, and K. L. Williams, "Direct measurement of sediment sound speed using SAMS in SW06," J. Acoust. Soc. Am.**124**, EL116–EL121 (2008).
[10]B. J. Kraft, I. Overeem, C. W. Holland, L. F. Pratson, J. P. M. Syvitski, and L. A. Mayer, "Stratigraphic model predictions of geoacoustic properties," IEEE J. Ocean. Eng. **31**, pp. 266–283 (2006).

Geoacoustic inversion on the New Jersey Margin: Along and across the shelf

Megan S. Ballard
Graduate Program in Acoustics, The Pennsylvania State University, State College, Pennsylvania 16801
msd200@psu.edu

Kyle M. Becker
Applied Research Laboratory, The Pennsylvania State University, State College, Pennsylvania 16803
kmbecker@psu.edu

Abstract: This paper presents results of a range-independent perturbative inverse approach applied to data from the Shallow Water Experiment 2006. The inversion technique is based on a linearized relationship between sound speed in the sediment and modal eigenvalues. Horizontal wave numbers were estimated from data collected from two distinct source/receiver tracks oriented along and across the shelf. The specific inversion algorithm used is based on qualitative regularization and uses known information about the environment to constrain the solution. Locations of the R reflector and other layering information are used as *a priori* information for the inversion to emphasize the layered structure of the sediment.

PACS numbers: 43.30.Pc, 43.30.Bp [WC]
Date Received: May 14, 2008 **Date Accepted:** June 13, 2008

1. Introduction

In the summer of 2006, the Office of Naval Research sponsored the Shallow Water Experiment 2006 (SW06) on the New Jersey Shelf area of the North Atlantic. This environment is known for the presence of nonlinear internal waves in the water column and high spatial variability of the seabed properties and has been a site of intense scientific interest.[1–3] Despite the variability of the seabed, there are prominent features that persist over large areas of the shelf; an example being the "R" reflector, a prominent shallow subsurface seismic reflector that varies in depth with location on the shelf.[4] SW06 is unique in the degree to which high resolution oceanographic measurements were made for support and interpretation of acoustic measurements and results. In this context, geoacoustic inversion methods are being developed that are sensitive to variability in the seabed, but robust to the dynamics of the water column. This work presents preliminary inversion results from SW06 for future comparison and synthesis with other inversion results in the same area as they become available.

In this work, perturbative inversion,[5] using modal wave numbers as input data, is used to determine compressional sound speed as a function of depth in the seabed. The approach used here differs from previous perturbative inversion work in that the locations of the R reflector and other layering information are used as *a priori* information in the inversion algorithm. This solution approach better approximates the layered structure of the sediment. Large scale spatial variability of the seabed is examined by estimating geoacoustic parameters using data collected from two distinct source/receiver tracks oriented along and across the shelf relative to a common receiver array location.

2. Description of experiment

The acoustic data collected during SW06 for this work were recorded on a 16-channel vertical line array (VLA) spanning the water column between 13.5 and 78 m depth. Broadcasting continuous tones at 50, 75, 125, and 175 Hz, a low-frequency acoustic source was towed at constant depth from the R/V Endeavor. The source was towed out and back along radials with respect to

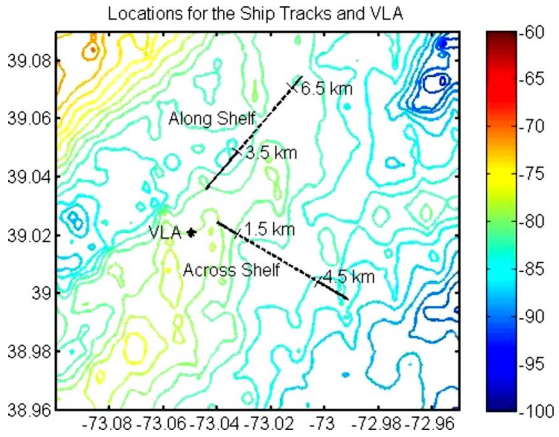


Fig. 1. (Color online) Contour map showing bathymetry with locations of the VLA and ship tracks overlaid. The segment of each ship track used for wave number estimation and inversion is shown by the dashed portion of the tracks. The range from the VLA at the beginning and end of this segment is listed on the figure.

the VLA, and repeated at speeds of 2–10 knots, to create synthetic apertures of ~5 km. Ship tracks, oriented along and across the shelf, and the VLA location are shown in Fig. 1. The flattest 3 km segment of each track, as measured by ship fathomer, was used for wave number estimation and inversion and is shown by the dashed portion of the tracks. Sound speed in the water column was determined using a combination of conductivity, temperature, and depth (CTD) data from sensors mounted on the VLA and 16 other nearby moorings,[6] along with a CTD chain towed by the source ship. Sound speed profiles used in the following analysis were determined by averaging measurements made during the acoustic data runs.

3. Summary of results

3.1 Wave number estimation

Auto regressive (AR) techniques[7] were used to estimate wave numbers from the 125 and 175 Hz pressure field data. Mode shapes, estimated by applying the large-argument Hankel transform[7] to data recorded on each receiver of the VLA, plotted versus depth are shown in Fig. 2 along with the AR spectrum at 125 Hz. Mode shapes aided in identifying modes, particularly mode 6, which had very low amplitude and appeared to split. Mode splitting has been observed in previous work[8] and, among other things, can be caused by range dependence or by the underlying sediment structure. The nature of the splitting observed here will be reported in future work by combining detailed sub-bottom profiling data with high-resolution wave number estimation methods.

Wave number values used as input into the inversion scheme were obtained by averaging over estimates obtained at each receiver depth. Doppler effects were removed by taking the mean of wave number estimates for incoming and outgoing tracks at the same speed.[9] To generate statistics for error analysis, standard deviations of the final estimates were also computed. Figure 3 shows normal distributions calculated from the mean and standard deviation of each mode number for the across shelf 125 Hz data. Low order modes have narrower distributions, reflecting greater signal to noise relative to higher order modes, as indicated in Fig. 3. Mode 6, which appeared split, has a much wider distribution.

3.2 Perturbative inversion

Sound speeds in the sediment for the along and across-shelf tracks were obtained using perturbative inversion with qualitative regularization (QR). Derived from Tikhonov regularization, QR chooses a solution which best fits some prior knowledge.[10] The solution is a piece-wise smooth profile with discontinuities at predetermined locations. Inversion results obtained using

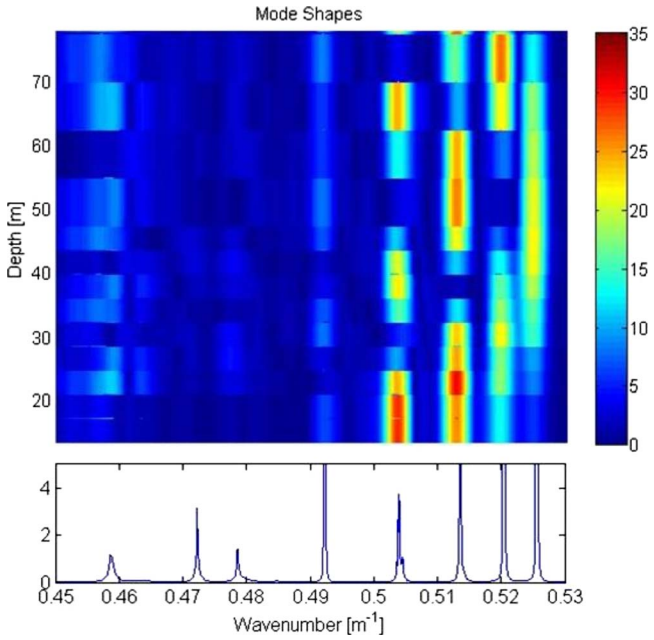


Fig. 2. (Color online) Mode shapes and the Green's spectrum estimated for the across shelf track. Mode 6 suffers from peak splitting and is labeled twice in the figure. Mode 8 is not shown in this figure.

wave numbers estimated from the 125 and 175 Hz pressure data are shown in Fig. 4 and are consistent with other experimental results obtained on the NJ shelf.[2,3] The two profiles are very similar; the difference in sound speed is less than 20 m/s at each depth indicating the overall structure of the sediment to be consistent over the area covered by the along and across shelf tracks. The depths at which discontinuities in the profiles occur were provided as *a priori* information for qualitative regularization. The locations were determined by an initial interpretation of sub-bottom profiling data taken during the SW06. The upper discontinuity was chosen to coincide with a transition from inhomogeneous clay to a finely layered structure at 10 m, and the second was the location of the R reflector at 18 m.[11] The relatively large jump in sound

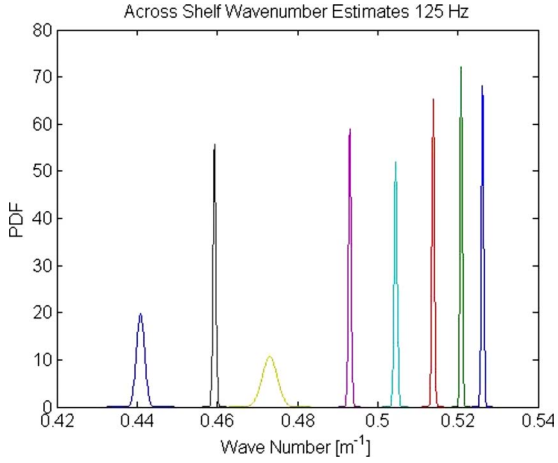


Fig. 3. (Color online) Distributions of the wave number estimates calculated from the mean and standard deviation of the estimates.

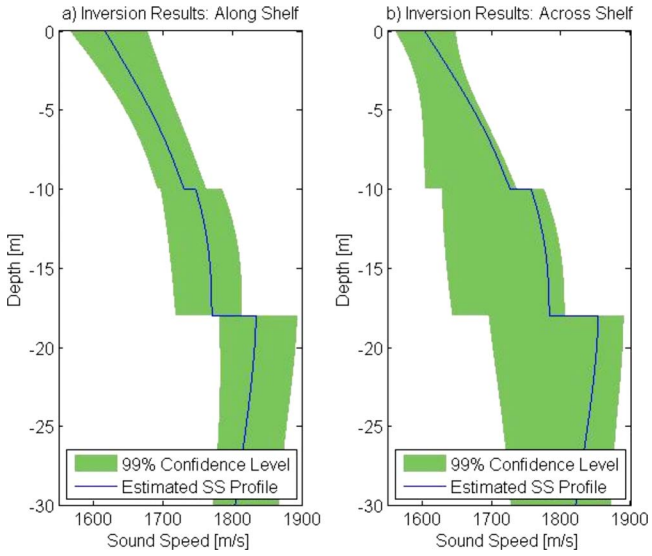


Fig. 4. (Color online) Inverted sediment sound speed profiles for (a) along shelf and (b) across shelf tracks. Best profiles are indicated by solid line with 95% confidence interval indicated by wide band.

speed at 18 m in the inversion results is consistent with the impedance contrast provided by the R reflector at that depth. Inversion result quality was evaluated by calculating the residual, defined as the 2 norm of the difference between wave numbers measured from data and wave numbers calculated from the model, at each frequency as given in Table 1. Residuals were calculated for all four frequencies and indicate a good model data match.

Estimates of the uncertainty of the inversion results were determined by Monte Carlo methods.[12] Realizations of noisy wave number data used in the inversion algorithm were generated using the measurement statistics shown in Fig. 3. Ninety five percent confidence intervals as a function of depth for the inversion results are indicated by the wide band in Fig. 4. The confidence intervals were formed under the assumption of normally distributed model parameters. This assumption works well for the along shelf inversion results. However, the across shelf inversion had model parameters with a bimodal distribution with the inversion result corresponding to the lesser of the two maxima. Consequently, the inversion result lies near the edge of the confidence interval.

4. Conclusions

Wave numbers for four different frequencies were estimated from the waveguide using an auto regressive estimator for two orthogonal tracks on the New Jersey shelf. Range independent inversions were performed according to the perturbative method using qualitative regularization. The resulting sediment sound speed profiles revealed similar gross characteristics of the seabed along and across the shelf in this area and were consistent with previously reported values for the New Jersey shelf. However, variability in the depth of the "R" reflector could not

Table 1. Residual calculated at all four frequencies using the inversion results.

Frequency	Along shelf	Across shelf
50	0.0006	0.0010
75	0.0022	0.0023
125	0.0015	0.0015
175	0.0016	0.0016

be resolved using the range-independent methods. Detailed range-dependent inversion methods should be applied to this and other data sets to obtain a three-dimensional characterization of the sediment in this region.

Acknowledgments

We would like to thank the scientific and engineering staff at the Woods Hole Oceanographic Institution for providing data acquisition, both acoustic and oceanographic, to support these experiments. Our appreciation is also extended to the crew of R/V Endeavor for towed source and CTD chain operations. M. S. Ballard was supported by a National Defense Science & Engineering Graduate Fellowship award. This work was also supported by an Office of Naval Research Special Research Awards in Underwater Acoustics Entry-Level Faculty Award, Grant No. N00014-04-1-0296.

References and links

[1]The SWARM Group, "An overview of the 1995 SWARM shallow water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500 (1997).

[2]S. D. Rajan, J. Doutt, and W. M. Carey, "Inversion for the compressional wave speed profile of the bottom from synthetic aperture experiments conducted in the Hudson Canyon Area," IEEE J. Ocean. Eng. **23**, 174–187 (1998).

[3]B. J. Kraft, I. Overeem, C. W. Holland, L. F. Pratson, J. P. M. Syvitski, and L. A. Mayer, "Stratigraphic model predictions of geoacoustic properties," IEEE J. Ocean. Eng. **20**, 321–336 (1995).

[4]T. A. Davies, J. A. Austin, Jr., M. B. Lagoe, and J. D. Milliman, "Late quaternary sedimentation off New Jersey: New results using 3-D seismic profiles and cores," Mar. Geol. **108**, 323–343 (1992).

[5]S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative Methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust. Soc. Am. **82**(3), 998–1017 (1987).

[6]A. E. Newhall, T. F. Duda, K. von der Heydt, J. D. Irish, J. N. Kemp, S. A. Lerner, S. P. Libertore, Y.-T. Lin, J. F. Lynch, A. R. Maffei, A. K. Morozov, A. Scmelev, C. J. Sellers, and W. E. Witzell, "Acoustic and oceanographic observations and configuration information for the WHOI moorings from the SW06 experiment," Woods Hole Oceanographic Institution Technical Report No. WHOI-2007-04 (2007).

[7]K. M. Becker and G. V. Frisk, "Evaluation of an autoregressive spectral estimator for modal wave number estimation in range-dependant shallow-water waveguides," J. Acoust. Soc. Am. **120**(3), 1423–1434 (2006).

[8]R. J. Cederberg, W. M. Carey, and W. L. Siegmann, "Modal analysis of geoacoustic influences on shallow-water propagation," IEEE J. Ocean. Eng. **22**, 237–244 (1997).

[9]K. M. Becker, "Accounting for bias in horizontal wavenumber estimates due to source motion," in *Acoustic Sensing Techniques for the Shallow Water Environment: Inversion Methods and Experiments*, edited by A. Caiti, N. R. Chapman, J.-P. Hermand, and S. M. Jesus (Springer, Dordrecht, 2006).

[10]L. Simcik and P. Linz, "Qualitative regularization: Resolving non-smooth solutions," Report No. CSE-94-12, Dept. of Computer Science, University of California Davis (1994).

[11]J. Goff, J. A. Austin, Jr., B. Christensen, and A. Turgut, "Chirp seismic reflection data on the New Jersey Middle and Outer Shelf: The geologic response to 40 000 years of sea level change," J. Acoust. Soc. Am. **122**(5), Pt. 2 2983 (2007).

[12]R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier, Oxford, 2005).

Geoacoustic inversion using combustive sound source signals

Gopu R. Potty and James H. Miller
Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02882
gopu@uri.edu, miller@egr.uri.edu

Preston S. Wilson
Applied Research Laboratory, University of Texas, P.O. Box 8029, Austin, Texas 78713-8029
pswilson@mail.utexas.edu

James F. Lynch and Arthur Newhall
Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts
jlynch@whoi.edu, anewhall@whoi.edu

Abstract: Combustive sound source (CSS) data collected on single hydrophone receiving units, in water depths ranging from 65 to 110 m, during the Shallow Water 2006 experiment clearly show modal dispersion effects and are suitable for modal geoacoustic inversions. CSS shots were set off at 26 m depth in 100 m of water. The inversions performed are based on an iterative scheme using dispersion-based short time Fourier transform in which each time-frequency tiling is adaptively rotated in the time-frequency plane, depending on the local wave dispersion. Results of the inversions are found to compare favorably to local core data.

PACS numbers: 43.30.Pc, 43.60.Hj, 43.60.Mn [WC]
Date Received: March 26, 2008 **Date Accepted:** April 10, 2008

1. Introduction

The "long range sediment tomography technique," developed by Potty *et al.* (2000), uses the arrival time dispersion of acoustic normal modes to estimate the compressional wave speeds of the ocean sediment. The technique is based on matching the observed modal arrival times with predicted ones for various "candidate models." The search for the model which produces the best match is carried out using a global optimization scheme. Time-frequency analysis using the short time Fourier transform (STFT) or the wavelet transform (WT) enables us to accurately observe the modal arrival times. Time-frequency analysis of combustive sound source (CSS) data collected during shallow water (SW06) shows fewer mode arrivals (3–5 modes) compared to explosive sources (7–9 modes; Potty *et al.*, 2000). These arrivals are in the frequency range 20–200 Hz and resolution of the modal arrivals is poor at the lower frequencies (20–100 Hz). Considering the fewer number of mode arrivals present in the data it is important to increase the quality of the available arrivals especially at lower frequencies. In order to achieve this we use a new adaptive approach which utilizes the local group velocity in the time-frequency analysis. An iterative scheme inputting acoustic data collected during the (SW06) experiment is used to estimate the compressional wave speeds of the sediments in the New Jersey shelf area. These results will be compared to other bottom property studies in the region.

2. Dispersion based short time Fourier transform (DSTFT)

Conventional time-frequency analysis techniques such as the STFT and the WT have often been used for the analysis of dispersive waves. However, their time-frequency tilings do not consider the dispersion effect explicitly. We have modified these approaches and used an adaptive time-frequency analysis method, originally proposed by Hong *et al.* (2005), whose time-frequency tilings depend on the dispersion characteristics of the signal to be analyzed. This method, called

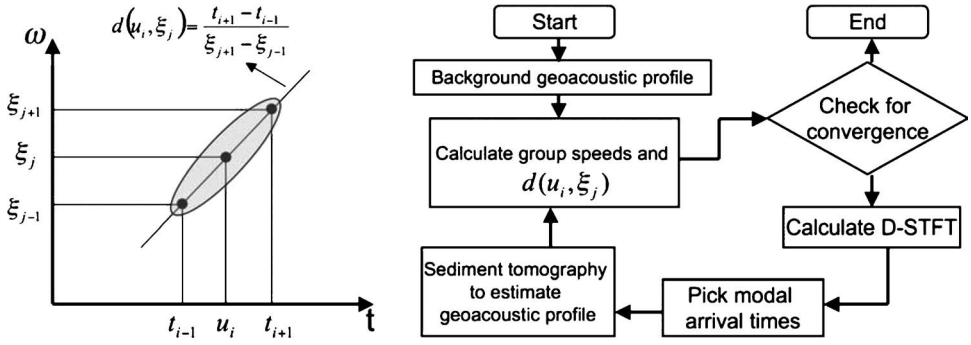


Fig. 1. Determination of the rotation parameter (left panel) and the steps involved in the inversion process using DSTFT (right panel)

the dispersion based short time Fourier transform (DSTFT), is performed by adaptively rotating each of the analysis atoms in the time-frequency plane with respect to the wave dispersion relationship.

For a square integrable function $f(t)$, the DSTFT at any time-frequency location (u,ξ) is defined as follows:

$$Df(u,\xi)=\int_{-\infty}^{\infty} f(t)\bar{g}_{(s,u,\xi,d)}(t)dt=\int_{-\infty}^{\infty} f(t)\left[\frac{1}{\sqrt{s}}g\left(\frac{t-u}{s}\right)\otimes(id)^{-1/2}e^{-i(t^2/2d)}\right]e^{-i\xi t}dt, \quad (1)$$

where $\bar{g}$ denotes the complex conjugate of the analysis window $g(t)$ which is given by

$$g_{(s,u,\xi,d)}(t)=\left[\frac{1}{\sqrt{s}}g\left(\frac{t-u}{s}\right)\otimes(id)^{-1/2}e^{-i(t^2/2d)}\right]e^{i\xi t}. \quad (2)$$

The symbol $\otimes$ denotes the convolution operator, parameter s determines the size of the window, and the parameter d determines the amount of rotation of the time-frequency tile in (u,ξ), such that

$$d=d(u,\xi)=\frac{\Delta u}{\Delta\xi}. \quad (3)$$

The group delay of the basis function of Eq. (2) in the time-frequency plane is given by Hong *et al.* (2005)

$$\tau(\omega)=\frac{d}{d\omega}\left[u(\omega-\xi)+\frac{d}{2}(\omega-\xi)^2\right]=u+d(\omega-\xi). \quad (4)$$

Equation (4) implies that the time-frequency tile in (u,ξ), using the basis function of Eq. (2), can be obtained by rotating or shearing the time-frequency tile of the standard STFT, using the parameter $d(u,\xi)$. Therefore if we choose $d(u,\xi)$ with respect to the local wave dispersion, the resulting time-frequency tiling will correspond to the entire wave dispersion behavior.

To use the DSTFT, the value of the parameter d has to be evaluated, based on the local dispersion behavior, at all the locations of interest (u,ξ) in the time-frequency plane. The procedure for the calculation of "d" is graphically illustrated in Fig. 1 (left panel). The parameter $d(u,\xi)$ can be written as

$$d(u_i,\xi_j)=\frac{\Delta t}{\Delta\omega}=\frac{t_{i+1}-t_{i-1}}{\xi_{j+1}-\xi_{j-1}}=\frac{\lfloor D/C_g(\xi_i+1)-D/C_g(\xi_i-1)\rfloor}{\xi_{j+1}-\xi_{j-1}}, \quad (5)$$

where D is the range and C_g is the group velocity at the location (u_i,ξ_j). To begin the inversion process we need to assume background values for the group speeds as a function of frequency corresponding to various modes. Based on these values we can initialize the inversion process by calculating the rotation parameter $d(u,\xi)$ using Eq. (5). Using the values of d at all time-frequency locations, the DSTFT can now be computed and the arrival times corresponding to individual modes can be picked as a function of frequency. These modal arrival times form the input data for our sediment tomography technique (Potty *et al.*, 2000). Output of this inversion provides values of the compressional wave inversion of the sediments. The group velocities of the normal modes are then recomputed based on these new sediment sound speed values and new estimates of the rotation parameter $d(u,\xi)$ are found using Eq. (5). Time-frequency analysis can now be performed again, using these new values of the rotation parameter, to recompute the modal arrival times. These iterative steps are repeated to achieve convergence of the group velocity estimates. Figure 1 (right panel) summarizes the steps involved in the inversion scheme described above.

3. Shallow water experiment 2006 (SW06) and combustive sound sources (CSS)

During the SW06 experiment, the combustive sound source (CSS) was deployed at 26 m depth from R/V Knorr in water depths of approximately 100 m. A combustible gas mixture of electrolytically derived hydrogen and oxygen was captured in the combustion chamber of the CSS and ignited with a spark. The ensuing burning produced expanding gases which in turn produced high intensity, low frequency acoustic pulses (Wilson *et al.*, 1995). The left panel in Fig. 2 shows the various stages of CSS generation for a conical shaped source. Cylindrical shaped sources were used in the SW06 experiment but the CSS generation is similar. These acoustic signals were received at the single hydrophone receiver units (SHRUs) deployed by the Woods Hole Oceanographic Institution. The time-frequency diagram of the received acoustic signal is shown in the right panel of Fig. 2. The first three acoustic mode arrivals are clearly seen and also modes 4 and 5 which are only partly present in the data. The SHRUs were deployed at water depths ranging from 65 to 110 m. The results presented here correspond to the CSS deployed at 39° 5.5174′ N and −73° 5.5816′ E and the receiver (SHRU) was located at 38° 57.6715′ N, −72° 54.8139′ E. The source to receiver range was 21.24 km. The bathymetry is range dependent and hence the acoustic propagation was modeled using adiabatic mode theory assuming six range sections. The sediment properties were assumed range independent and the depth variation of the sediments was modeled using sediment layers of unknown thickness over an acoustic basement.

4. Iterative scheme to estimate the compressional wave speed

A geoacoustic model based on historic data from the experimental location (Jiang *et al.*, 2007) was used in the beginning to initialize the parameter $d(u,\xi)$. Group velocities corresponding to each normal mode were used to compute $d(u,\xi)$ which were then assigned to time-frequency locations in the neighborhood of the normal mode. This is the key step in the time-frequency analysis and the inversion process—the modal group speeds as a function of frequency need to be mapped correctly into the time-frequency plane. The modal arrival times are thus chosen from the DSTFT and the tomography inversion is carried out using these values. A new estimate of the modal group speeds is made and the steps are then repeated to get a new estimate until the group speed values converge.

Figure 3 (left panel) shows the compressional wave speeds in the top 40 m of the sediments estimated by the inversion scheme. The dashed line is the geoacoustic model proposed by Jiang *et al.* (2007). They suggested this model for the SW06 experimental area based on chirp sonar inversions, grab samples, *in situ* probes, and shallow core measurements. The middle panel in Fig. 3 shows the standard deviations associated with these compressional wave

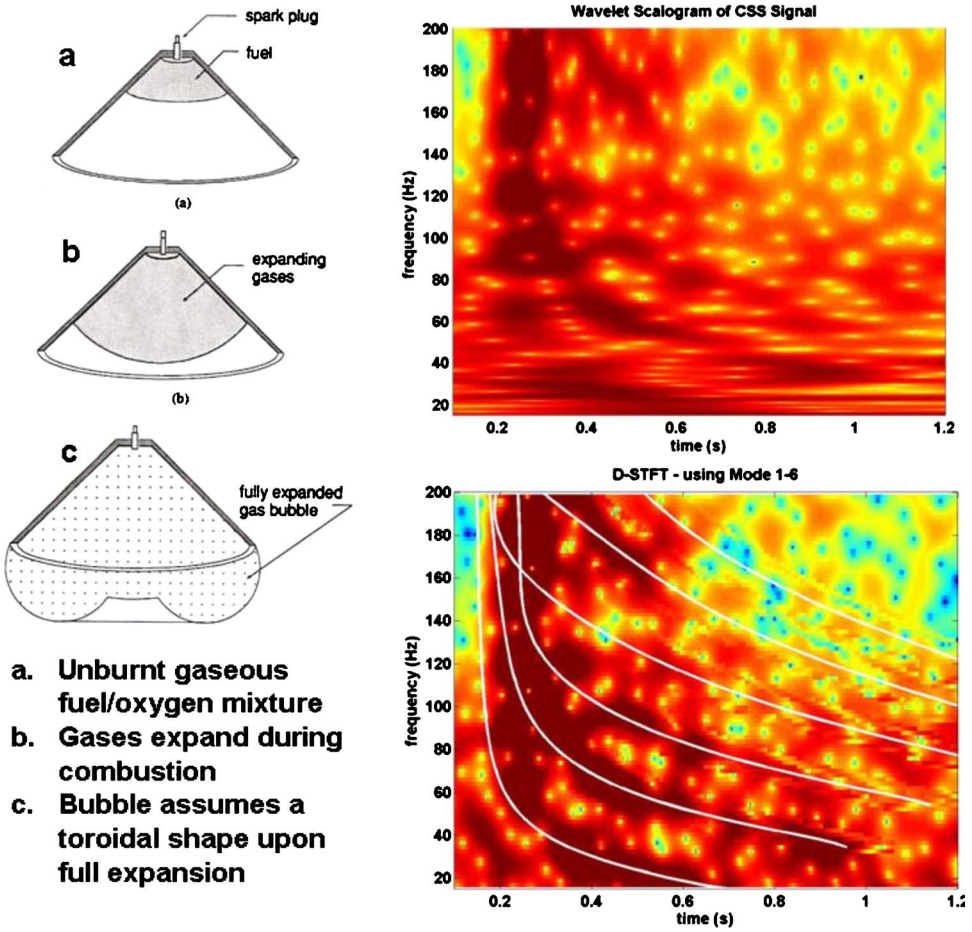


Fig. 2. (Color online). Left panel shows cross section of the CSS combustion chamber. The right panel shows the time-frequency diagram calculated using the wavelet (top) and DSTFT (bottom) methods. The continuous lines in the DSTFT diagram represent theoretical modal arrival times computed using the inversion results.

speed inversions. The standard deviations are of the order of 20 m/s except at 20 m depth where the uncertainty is higher. Compressional wave speeds in the top 15 m of the sediment are compared (right panel in Fig. 3) with data from the AHC-800 core (Norfjord, 2005) which was close to the source-receiver path. The AHC-800 data show the type of sediments and the sediment acoustic velocities in various layers. The sediments in the top 15 m are generally sandy in nature and interbedded with mud and shells. The inversion captures the trend present in the core data very well, though the average values of the compressional wave speeds are lower than the core data (but higher than the model of Jiang *et al.*). The maximum difference between the inversion and the core data as well as the model of Jiang *et al.* occurs in the near surface sediments (0–3 m). This could be due to the fact that only the lower order modes (in the frequency range 20–200 Hz) were used for the inversion. It should be also be noted that the inversions correspond to much lower frequencies compared to the frequencies at which the cores are usually logged (~20 kHz). This can cause differences in values of compressional wave speeds obtained from cores and inversion if frequency dispersion effects are present. The time-frequency diagram was calculated using the DSTFT with input geoacoustic values corresponding to the inversion is shown in Fig. 2. We can note that the modes are now better resolved in time at lower frequencies, which is particularly evident in the case of mode 1. The solid lines in

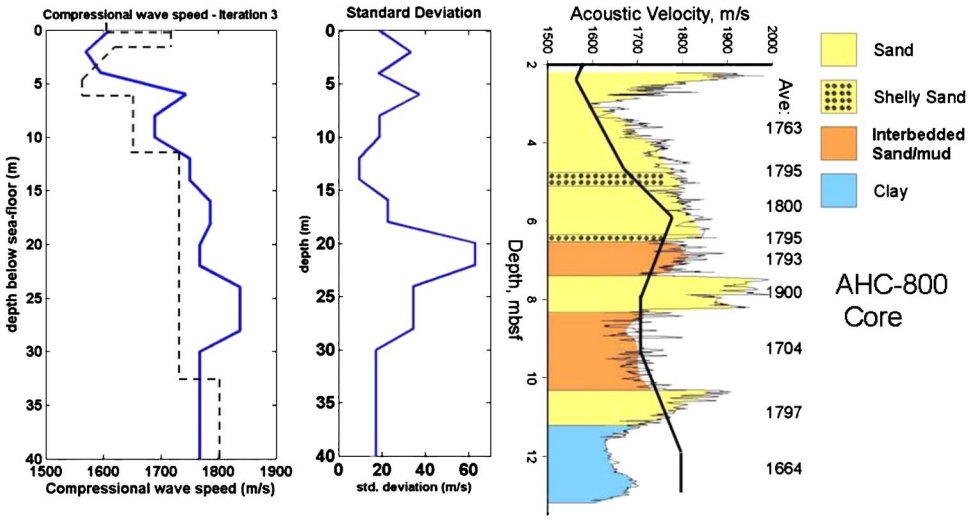


Fig. 3. (Color online) The inversion results for the top 40 m of the sediment (left panel) and the standard deviation (middle panel). The dashed line in the left panel is the Jiang-Chapman geoacoustic model (2007). The right panel shows the comparison of the inversion with data from AHC-800 core. The continuous black line in the right panel is the inversion result.

the figure correspond to the theoretical arrival times calculated based on the inversion. They agree well in most of the frequency bands, except for modes 1 and 3 in the 140–200 Hz band. Efforts are currently under way to repeat the inversions using data from other sources/receivers to explore the spatial variability of geoacoustic properties. Inversions to estimate the compressional wave attenuation will also be carried out using the CSS data.

5. Conclusions

Geoacoustic inversions were successfully performed using CSS data collected during the SW06 Experiment. A new time-frequency analysis approach is introduced which uses the local group velocity values to rotate the time-frequency tiles. The time-frequency analysis is linked to the inversion which, after each iteration, provides new estimates of the modal group velocity values based on the inversions (compressional wave speeds). The results of the inversion compare well with published core data from the region.

Acknowledgments

This study was supported by Office of Naval Research, Ocean Acoustics Program (Code 321 OA). The authors would also like to thank Professor Yoon Young Kim (Seoul National University) for providing the software for the dispersion based short-time Fourier transform analysis and helping with its implementation.

References and links

Hong, J.-C., Sun, K. H., and Kim, Y. Y. (**2005**). "Dispersion based short-time Fourier transform applied to dispersive wave analysis," J. Acoust. Soc. Am. **117**(5), 2949–2960.
Jiang, Y.-M., Chapman, N. R., and Badiey, M. (**2007**). "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data," J. Acoust. Soc. Am. **121**(4), 1879–1894.
Norfjord, S. (**2005**). *Late Quaternary geologic history of New Jersey middle and outer continental shelf*, PhD thesis, University of Texas at Austin, p. 202.
Potty, G. R., Miller, J. H., Lynch, J. F., and Smith, K. B. (**2000**). "Tomographic Inversion for sediment parameters in shallow water," J. Acoust. Soc. Am. **108**(3), 973–986.
Wilson, P. S., Ellzey, J. L., and Muir, T. G. (**1995**). "Experimental investigation of the Combustive Sound Source," IEEE J. Ocean. Eng. **20**(4), 311–320.

Seabed acoustics of a sand ridge on the New Jersey continental shelf

D. P. Knobles and P. S. Wilson
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713
knobles@arlut.utexas.edu, pwilson@mail.utexas.edu

J. A. Goff
Institute of Geophysics, The University of Texas at Austin, Austin, Texas 78713
goff@ig.utexas.edu

S. E. Cho
Department of Electrical and Computer Engineering, University of California, San Diego, California 92093
cho.steve.e@gmail.com

Abstract: Acoustic measurements were made on a sand ridge on the New Jersey continental shelf. Data collected on two L arrays separated by 20 km from a single multi-frequency tow suggest small horizontal environmental variability. Values for the sound speed structure of the seabed are extracted by first applying a geo-acoustic inversion method to broadband and narrowband acoustic data from short-range sources. Then, a parabolic equation algorithm is used to properly include the bathymetry and sub-bottom layering. Finally, the frequency dependence of the seabed attenuation is inferred by optimizing the model fit to long-range transmission loss data in the 50–3000 Hz band.

PACS numbers: 43.30.Ma, 43.30.Pc [WC]
Date Received: March 3, 2008 **Date Accepted:** May 16, 2008

1. Introduction

The frequency dependence of attenuation at low frequencies in sandy marine sediments remains an issue in the ocean acoustics community. In principle, inversion of acoustic measurements made in the water column can yield estimates of the sound speed and attenuation in the seabed. Sub-bottom layering, e.g. obtained with chirp sonar, and geophysical seabed properties, such as grain size distribution, constrain the inverse problem and aid the interpretation of results. Water column sound speed profile and bathymetry measurements provide additional required information for the inverse problem. Over short propagation ranges, the coherent structure of the acoustic field over a large frequency band assuming nominal values for the attenuation gives, along with geophysical ground truth, first-order estimates of the sediment sound speed. Then, the attenuation can be extracted by modeling long-range data with a range-dependent propagation model. This approach affects a partial decoupling of the sound speed and attenuation within a multi-dimensional cost function parameter space. The prospect for success with such an approach is enhanced if the depth dependence of the geo-acoustic structure is approximately uniform with range, and the remainder of the unknown range inhomogeneities are weak.

The frequency dependence of seabed attenuation was addressed as part of a shallow water acoustic experiment on the New Jersey continental shelf. For this purpose two L arrays were positioned about 20 km apart on a ridge composed of an approximately uniform coarse sand to quantify and mitigate the range inhomogeneities along an acoustic propagation track. This letter reports on the inferred experimental values of the attenuation on the sand ridge in the 50–3000 Hz band.

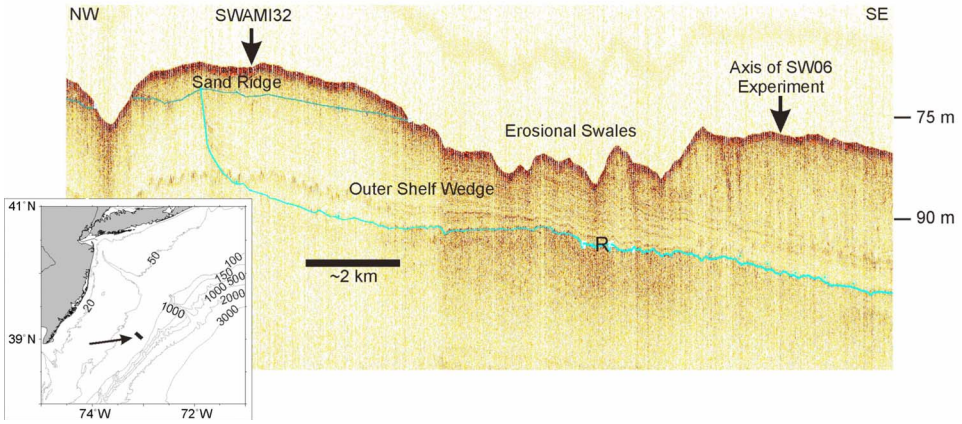


Fig. 1. (Color online) Sub-bottom chirp reflection profile of the seabed perpendicular to bathymetry isobaths and position of Array 1 and location of concentrated measurements during New Jersey continental shelf experiment.

2. Description of experiment and variability measurement along propagation track

Between 24 August and 06 September 2006 two arrays recorded both narrowband and broadband acoustic signals on the New Jersey continental shelf. The positions of Array 1 (a 32-element L array) and Array 2 (a 52-element L array), were 39.06° N, 73.13° W and 39.2° N, 72.97° W, respectively. Both arrays contained a vertical line array that spanned the water column 10 m beneath the water surface to the seabed and a horizontal line array about 270 m in length positioned on the seabed. The bandwidths of Array 1 and Array 2 were about 3000 and 1000 Hz, respectively.

Figure 1 shows a seismic reflection profile derived from chirp data (2–15 kHz) collected during the summer of 2006 just prior to the New Jersey shelf experiment. The positions of Array 1 and the track axis that formed the center of the experimental area are included on the image. By design Array 1 lies in the center of a sand ridge that extends into the figure about 20 km to the Array 2 position. The sand ridge was previously imaged with Huntec seismic data by Davies *et al.*[1] (their Fig. 5), and sampled by vibracoring (their Fig. 11). The thickness of the sand layer on the ridge is 3–5 m and the width of the ridge is about 6 km. Grain size distributions taken on the sand ridge have been reported by Goff *et al.*[2] On the propagation track between the two arrays sediment grain sizes lie between 0.8 and 1.5 ϕ, placing them in the medium to coarse sand category. Beneath the sand layer is the outer shelf wedge, a combination of fine sands, clays, and silt. The main axis of the New Jersey shelf experiment lies on the outer shelf wedge and the site of previous acoustic experiments in 1988 and 1993[3] was near the boundary of the sand ridge and the outer shelf wedge. In Fig. 1 to the right of this boundary a change in the level of sub-bottom reflections is observed. On the sand ridge the impedance at the water sediment interface is higher than on the outer shelf wedge, as indicated by the increased scattering at least 25 m down into the outer shelf wedge, whereas the scattered energy from the sub-bottom layering is significantly lower on the sand ridge. This implies that the seabed on the sand ridge consists of coarse sand, which produces a larger reflection coefficient compared to a sand-clay-silt mixture on the outer shelf wedge.

On an acoustic propagation track between the two arrays both narrowband and broadband sources were deployed. A sub-bottom layering track of the type in Fig. 1 coincides approximately with the propagation track and, with the measured bathymetry and sound speed profiles, provides input for a range-dependent propagation model. The narrowband data processed in the present study include signals recorded on both arrays from a towed J-15-1 source in the form of CW tones at 53, 103, 203, 253, 303, 403, 503, 703, and 953 Hz. Array 1 recorded mid-frequency data from a towed ITC 2015 source that produced tonals of 1153, 1503, 2003,

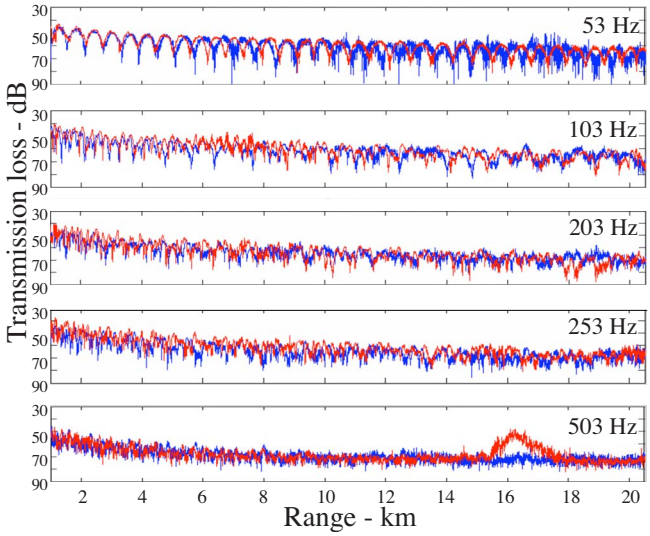


Fig. 2. (Color online) Experimental measurement of horizontal variability along propagation track. Array 1 (blue online) and Array 2 (red online).

2503, and 2953 Hz. The broadband data processed in the current analysis include the time series recorded at Array 1 generated from a combustive sound source (CSS) event.

The positioning of the two arrays permits an experimental determination of range inhomogeneity along the acoustic propagation track. Figure 2 shows measured transmission loss (TL) at the two arrays derived from the same J-15-1 tow between the two arrays. The levels and, to a significant degree, the coherent structure are receiver independent, suggesting that, while there may be time and range inhomogeneities, some of which cannot be accounted for in a propagation model due to their unknown origin, they are weak enough to permit the seabed attenuation to be inferred from long-range TL data.

3. Analysis

Inversion methods were applied to (1) a CSS event about 4.7 km from Array 1 and (2) low-frequency tow data received on both arrays from the main propagation track. The source time series for the CSS was recorded separately on a hydrophone 2 m from the source. In the broadband inversion approach simulated annealing is used with a cross-correlation cost function previously applied to impulsive data.[4–6] The narrowband inversion approach uses a cost function that sums coherently model-data cross spectrum pairs over frequency, pairs of subaperatures, and time.[5] For the CSS event, the frequency band was 35.5–264.4 Hz with $\Delta f = 1.144$ Hz for a total of 201 frequencies. The narrowband inversions over source-range spatial scales of about 7 km used frequencies 53, 103, 203, and 253 Hz.

Figure 3(a) shows a comparison of measured (at Array 1) and modeled (postinversion) CSS time series in the 50–3000 Hz band. Figure 3(b) shows a comparison of measured (at Array 2) and modeled TL for 53, 103, 203, and 253 Hz. The modeled time series and TL were computed using the inverted sound speeds in Tables 1 and 2. The cost function minimum for the sound speed ratio at the water sediment interface at both Array 1 and Array 2 was about 1.103, which gives a sound speed of about 1650 m/s. For the Array 1 location both the CSS and low-frequency tow data inversions (not shown here) reveal a sound speed ratio of less than unity between the first and second sediment layer, which is consistent with the geophysical data previously described in Sec. 2. For the Array 2 location the cost function minima of the sound speed ratio between the first and second sediment layer is greater than unity, which suggests the presence of consolidated sands beneath the sand layer.

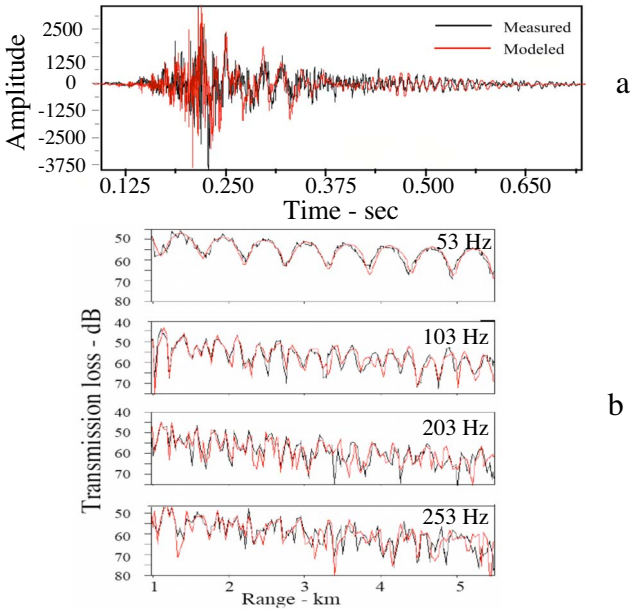


Fig. 3. (Color online) (a) Comparison of measured CSS received time series on the Array 1 array with postinversion modeled time series over 10–3000 Hz band and (b) comparison of postinversion modeled (red online) and measured (white online) TL on the Array 2 array.

The results of the geo-acoustic inversions performed at the two arrays, the measured bathymetry, the measured sub-bottom layering along the propagation track, and measured sound speed profiles near the two arrays during the time of the multi-frequency source tows were used to construct an approximate range-dependent waveguide. Absorption in the water column is included using a standard Thorp attenuation relationship. Figures 4(a)–4(c) show comparisons of modeled TL using a parabolic equation approach developed by Collins[7] and measured TL from the 50–3000 Hz band over ranges of 15–20 km. The attenuation in the first sediment layer has been adjusted as a function of frequency to optimize the fit with the measured data. Matching both the measured TL levels and overall range modal interference structure were important considerations in estimating the attenuation. Specifically as the range increases the modal interference pattern becomes less complicated (mode stripping). The character of the modal interference pattern evolution with range is therefore strongly dependent on the attenuation. Thus, in addition to the observed average trend of the TL with range, the evolution (qualitative) of the multimodal interference pattern over large range scales assisted in inferring the attenuation from the measured data.

Figure 4(d) shows the inferred attenuation values for the first sediment layer as a function of frequency. For purposes of comparison Fig. 4(d) includes several predictions of the Biot theory.[8,9] The specific Biot calculations are based on the formulation of Stoll and Kan.[10] The

Table 1. Sediment sound speed structure from CSS and multi-frequency inversions at Array 1.

Depth-m	Sound speed-m/s
0	1650
3.2	1653
3.2	1580
23	1600
23	1850

Table 2. Sediment sound speed structure from multi-frequency inversions at Array 2.

Depth-m	Sound speed-m/s
0	1651.9
3.02	1660.8
3.02	1809.8
34.4	1811.9
34.4	2036.9

solid back curve is the prediction using the Biot parameters listed in Table I of Ref. 11, which were derived from measurements made during the Sediment Acoustics Experiment in 1999 (SAX99). The solid (blue online) curve uses the SAX99 parameters but with a lower bound estimate (β=0.363) for the porosity at the New Jersey shelf location.[2] The two dotted lines use the SAX99 parameters but with upper and lower bound estimates for the permeability, $\kappa=1.0\times10-10$ m^2 and $\kappa=1.0\times10-11$ m^2, respectively. Permeability was not measured at the New Jersey shelf site, but these estimates were derived from sediment physical models that relate permeability to porosity and grain size.[12] Core samples for the New Jersey site[2] yielded a mean grain size of ϕ=1. For frequencies f below about 2 kHz the frequency dependence of the inferred attenuation is proportional to approximately $f^{1.85}$. These results are consistent with other studies that have inferred a nonlinear frequency dependence of the attenuation from analyses of forward propagation data.[3,13–16] While the data are not inconsistent with the Biot predictions above 2 kHz, one cannot draw any definitive conclusions concerning the frequency exponent because of the sparseness of data.

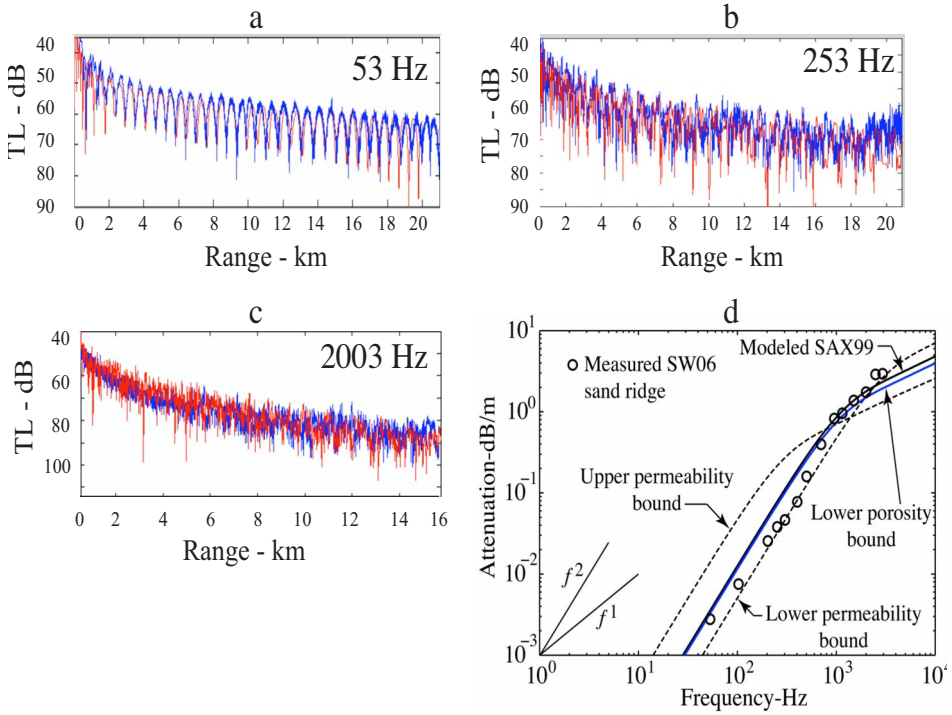


Fig. 4. (Color online) (a–c) Example comparisons of optimized modeled TL (red online) with measured TL (blue online) and (d) inferred attenuation versus frequency for first sediment layer.

4. Summary

Both broadband and narrowband acoustic measurements in the 50–3000 Hz band made on the New Jersey continental shelf were analyzed for the information they contain on the frequency dependence of the attenuation in the seabed. The array positions and propagation track were on a sand ridge. Geoacoustic inversions revealed sound speeds of the surface sand layer between 1650 and 1652 m/s, consistent with the existing geophysical data. For frequencies below about 2 kHz the attenuation varied approximately as $f^{1.85}$, consistent with predictions made by the Biot theory. While the inferred attenuation is consistent with predictions made by the Biot theory above 2 kHz, the data were too sparse to draw definitive conclusions concerning the frequency exponent of the attenuation.

Acknowledgments

The U.S. Office of Naval Research Code No. 321 OA supported this work. The first author gratefully acknowledges Dr. William Hodgkiss of the Marine Physical Laboratory, Scripps Institution of Oceanography, who provided an ITC2015 transducer source for the collection of acoustic data above 1 kHz on Array 1. Finally, the first author wishes to acknowledge helpful comments provided by an anonymous referee.

References and links

[1]T. A. Davies, J. A. Austin, Jr., M. B. Lagoe, and J. D. Milliman, "Late quaternary sedimentation off New Jersey: New results using 3-D seismic profiles and cores," Mar. Geol. **108**, 323–343 (1992).

[2]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).

[3]S. M. Dediu, W. L. Siegmann, and W. M. Carey, "Statistical analysis of sound transmission results obtained on the New Jersey continental shelf," J. Acoust. Soc. Am. **122**, EL23–EL28 (2007).

[4]D. P. Knobles, R. A. Koch, L. A. Thompson, and K. C. Focke, "Sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**, 205–222 (2003).

[5]R. A. Koch and D. P. Knobles, "Geo-acoustic inversion from surface ships," J. Acoust. Soc. Am. **117**, 626–637 (2005).

[6]D. P. Knobles, T. W. Yudichak, R. A. Koch, P. G. Cable, J. H. Miller, and G. Potty, "Inferences of seabed attenuation from distributed acoustic measurements in the East China Sea," IEEE J. Ocean. Eng. **31**, 129–144 (2006).

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

[8]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **26**, 168–178 (1956).

[9]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).

[10]R. D. Stoll and T. K. Kan, "Reflection of acoustic waves at a water sediment interface," J. Acoust. Soc. Am. **70**, 149–156 (1981).

[11]K. L. Williams, D. R. Jackson, E. I. Thorsos, D. J. 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).

[12]S. G. Schock, "A method for estimating the physical and acoustic properties of the sea bed using chirp sonar data," IEEE J. Ocean. Eng. **29**, 1200–2017 (2004).

[13]F. Ingenito, "Measurements of mode attenuation coefficients in shallow water," J. Acoust. Soc. Am. **53**, 858–863 (1973).

[14]J. X. Zhou, "Normal mode measurements and remote sensing of sea-bottom sound speed and attenuation in shallow water," J. Acoust. Soc. Am. **78**, 1003–1009 (1985).

[15]I. Rozenfeld, W. M. Carey, P. G. Cable, and W. L. Siegmann, "Modeling and analysis of sound transmission in the Strait of Korea," IEEE J. Ocean. Eng. **26**, 809–819 (2001).

[16]J. D. Holmes, W. M. Carey, S. M. Dediu, and W. L. Siegman, "Nonlinear frequency-dependence attenuation in sandy sediments," J. Acoust. Soc. Am. **121**, EL218–EL222 (2007).

Analysis of wind-driven ambient noise in a shallow water environment with a sandy seabed

D. P. Knobles and S. M. Joshi
Applied Research Laboratories, The University of Texas at Austin, Austin Texas 78758
knobles@arlut.utexas.edu, joshi@arlut.utexas.edu

R. D. Gaul
BlueSea Corporation, Houston, Texas 77014
rgaul@blueseacorp.net

H. C. Graber and N. J. Williams
University of Miami, Miami, Florida 33149
hgraber@rsmas.Miami.edu, nwilliams@rsmas.miami.edu

Abstract: On the New Jersey continental shelf ambient sound levels were recorded during tropical storm Ernesto that produced wind speeds up to 40 knots in early September 2006. The seabed at the position of the acoustic measurements can be approximately described as coarse sand. Differences between the ambient noise levels for the New Jersey shelf measurements and deep water reference measurements are modeled using both normal mode and ray methods. The analysis is consistent with a nonlinear frequency dependent seabed attenuation for the New Jersey site.

PACS numbers: 43.30.Nb, 43.30.Ma [WC]
Date Received: March 7, 2008 **Date Accepted:** May 16, 2008

1. Introduction

The issue of the nature of wind-generated ambient noise (AN) in ocean environments is a topic that has received considerable attention in the history of underwater acoustics. Kundsen[1] and Wenz,[2] using acoustic data collected primarily in deep water, reported average third octave noise levels versus frequency and wind speed. At frequencies above 500 Hz their averaged levels are in close agreement. Supplemented by other deep-water measurements, these studies are the basis for wind-generated deep-water AN level *reference* curves.[3] The term "deep-water" denotes the condition of AN dominated by signals that propagate along the direct and surface-reflected paths within the Reliable Acoustic Path range. Because of the dominance of high-angle ray paths (or modes), the wind-driven AN levels in deep water are approximately independent of the hydrophone receiver depth. The deep-water reference curves apply to the great majority of deep-water ocean environments and receiver geometries.

Proximity of the seabed to the sea surface alters propagation of wind-generated sound in shallow water compared to deep water environments. Under downward refracting conditions the dominant ray paths become bottom interacting within short ranges so the details of the seabed structure play an important role in the nature of the AN. Ingenito and Wolf[4] studied the site variability of wind-generated AN in shallow water areas (on the order of 10 dB), concluding that the variation was due to environmental differences. They introduced a methodology that used the known properties of waveguides to model site-dependent environmental effects. The results suggested that such an approach would give a more accurate prediction of AN levels in shallow water than obtained from reference AN curves. This paper implements a variant of the Ingenito and Wolf approach. AN levels recorded on the New Jersey continental shelf are modeled using previously determined properties of the seabed from geophysical and geoacoustic studies. The analysis demonstrates that the frequency dependence of the AN levels is sensitive to certain features of the seabed geoacoustics including the frequency dependence of the seabed

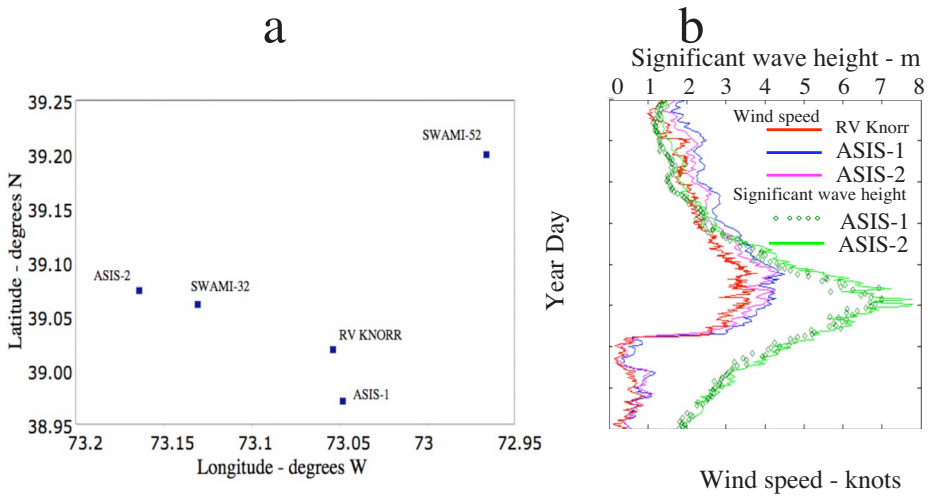


Fig. 1. (Color online) (a) Positions of acoustic arrays, surface buoys, and RN Knorr during Ernesto, (b) wind speed and significant wave height measurements during Ernesto.

attenuation.

2. Measurements

A segment of the Shallow Water Experiment 2006 was conducted on the New Jersey Continental Shelf (NJCS) in about 70–75 m of water from 24 August to 06 September 2006.[5] The main objective of the experiment was to increase fundamental understanding of sound interaction with the seabed in the 50–3000 Hz band. During 01 September to 02 September 2006 the center of tropical storm Ernesto passed through western New Jersey and on into Pennsylvania. Although the New Jersey shelf experimental area was about 150 miles from the center of the storm, wind speeds and significant wave heights of up to 40 knots and 24 feet, respectively, were recorded. An analysis of geophysical data and broadband and narrowband acoustic data have determined that a geoacoustic description valid for the New Jersey sites discussed in this paper is a top sediment layer composed of sand with a sound speed on the order of 1650–1700 m/s and an attenuation that varies nonlinearly with frequency.[5] The measured AN levels from the storm provide a unique opportunity to apply the Ingenito and Wolf concept because the New Jersey continental shelf environmental properties are established reasonably well.

Figure 1(a) shows the locations of two Shallow Water Acoustic Measurement Instrumentation (SWAMI) arrays and two Air Sea Interaction Spar (ASIS) buoys. SWAMI-32 and SWAMI-52 were configured as a 32-element L array and a 52-element L array, respectively. The ASIS buoys measured a variety of oceanographic and meteorological quantities including wind speed and direction and wave height and direction. Also shown in Fig. 1(a) is the nominal position of the RV Knorr whose anemometer measured wind speed during tropical storm Ernesto. Figure 1(b) shows the wind speed measurements collected by the RV Knorr and the two ASIS buoys. Also shown is the significant wave height measurements made by the two buoys. The wind speed values shown for the buoys and the RV Knorr have been converted to the standard reference value of 10 m. One observes moderate spatial variability of wind speed at the three locations prior to the peak of the storm and a significantly reduced variability after the peak. The reason for the lower wind speeds measured by the Knorr is not known. The superstructure of the ship was not taken into account in converting the Knorr measurements to the 10 m reference values.

Figure 2(a) shows an omni level spectrogram from a hydrophone on the horizontal segment of the SWAMI-32 array. The wind speed, measured on the ASIS-2 buoy, is superim-

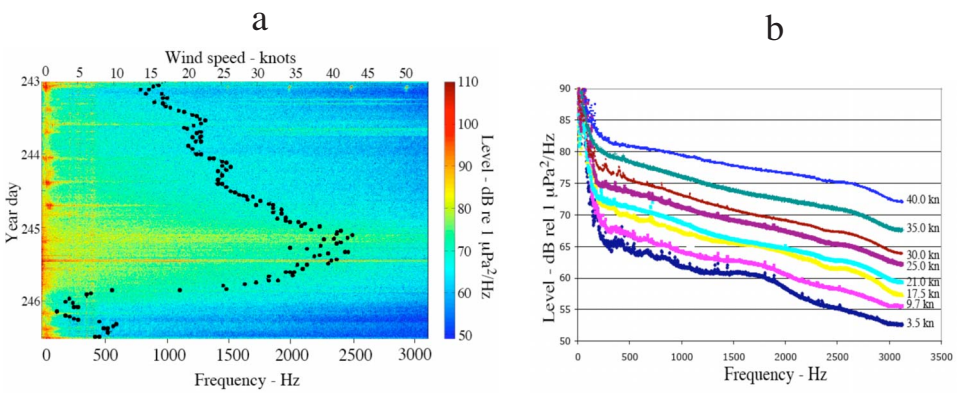


Fig. 2. (Color online) AN measurements at SWAMI-32.

posed on the spectrogram. The AN levels are clearly well correlated with the wind speed. Figure 2(b) shows the averaged AN levels deduced from measurements on the SWAMI-32 array for wind speeds from about 3–40 knots. The methodology used in this study for extracting the average levels as a function wind speed and frequency has been previously described.[6] The wind-generated region includes frequencies above about 400 Hz for wind speeds above about 3 knots. Below 400 Hz it becomes increasingly difficult to separate the distant shipping and wind components of AN in shallow water. It was observed (not shown here) that the averaged AN levels at the SWAMI-52 location in the 200–1000 Hz band were within about 1 dB of the AN levels measured at the SWAMI-32 site.

3. Noise Model

One approach to model AN levels is a method such as proposed by Kuperman and Ingenito that solves for the absolute received levels due to an agitated sea surface driven by wind.[7] This approach requires knowledge of the wind-generated source levels as a function of wind speed and frequency. If one assumes that for a given wind speed and frequency the wind-generated source levels are site independent, then an approach can be developed that does not require detailed knowledge of the source levels. The idea used in the current analysis is to compare observed and modeled differences in wind-driven AN levels between the New Jersey continental shelf environment and a reference deep water environment. The reference approach replaces the need to know source levels with the need to know the acoustics in a typical deep-water environment.

The difference in AN levels measured at two locations for a specific frequency, f, and wind speed, V, is

$$\Delta \mathrm{AN}_{12}(f,V) = \mathrm{AN}_1(f,V) - \mathrm{AN}_2(f,V). \tag{1}$$

Assuming that the source levels for a specific wind speed are independent of site, $\Delta \mathrm{AN}_{12}(f,V)$ is approximately independent of V. Further define for site j

$$L_j(z_0,z_j,f) = 10\ \mathrm{Log}_{10}\left(2\pi \int_0^{R_j} I(r) r dr\right), \tag{2}$$

where

$$I(r,z_0,z_j,f) = 10^{-TL_j(r,z_0,z_j,f)/10}. \tag{3}$$

L_j represents the AN level at site j for frequency f at depth z_j due to a 0 dB source with strength per unit area distributed in the x-y plane at depth z_0. TL_j represents the modeled TL and R_j is the maximum range of integration. R_j is chosen such that the integral in Eq. (2) has converged to within less than 0.1 dB. It is assumed that the waveguide possesses azimuthal symmetry and

that the source level is distributed uniformly over the area of integration. Finally define

$$\Delta L_{12}(f) = L_1(z_0,z_1,f) - L_2(z_0,z_2,f). \quad (4)$$

The main thesis of this paper is that

$$\Delta \mathrm{AN}_{12}(f) = \Delta L_{12}(f). \quad (5)$$

For this analysis, an environment studied by Gaul *et al.*[6] in the NE Pacific for a data collection called Church Opal (CO) is considered a representative or canonical waveguide for deep water measurements, as summarized by Ross.[3] The rationale is that for a typical deep-water environment that possesses depth excess beneath the deep sound channel, the wind-driven AN is dominated by the high-angle modes (or rays) that describe acoustic propagation out to the Reliable Acoustic Path range. The high-angle character of the AN is consistent with the observation that the wind-driven AN in a deep water environment is approximately depth independent.[8,9] For such environments the details of the seabed play only a small role in the high-angle (continuum) portion of the acoustic field. Given the pertinent environmental properties, the differences in AN levels at the two sites were calculated as a function of frequency using high angle propagation algorithms.[10–12] Specifically, for the NJCS environment a high angle normal mode approach[10,11] was used for all frequencies. For the CO environment the normal mode model was also used for frequencies less than 1500 Hz. It was observed that the normal mode model and a ray approach[12] give essentially identical results for the CO environment for frequencies below 1500 Hz. Above 1500 Hz the normal mode algorithm as implemented became highly inefficient and the results for the CO environment shown in this paper above 1500 Hz were computed using the ray approach.

4. Results

The properties of the New Jersey and CO environments have been previously described.[5,6] The water depth for the CO waveguide is about 4883 m. The SSP has its sound channel axis at about 600 m and the critical depth at about 4500 m. Absorption in the water column is included using a standard Thorp attenuation relationship. The seabed structure is described as a 10 m clay sediment over basalt. To a very good approximation the CO waveguide is horizontally stratified. The receiver depth used in the calculation for the CO environment was 4000 m. The seabed structure at the SWAMI-32 location is approximately described as a 3 m sand layer over a lower speed layer of a sand-clay mixture over a high speed consolidated sand layer. The sound speed of the first layer is about 1650 m/s. An important aspect of the geoacoustic structure of the New Jersey shelf site is that the attenuation, A(dB/m), has a nonlinear dependence with frequency. For frequencies below about 2 kHz the attenuation varied approximately as $f^{1.85}$ and for frequencies greater than about 2 kHz the attenuation varied approximately as $f^{1.4}$. The SSP at the New Jersey site employed for the calculation of Eq. (2) was a typical profile measured just prior to the arrival of Ernesto.

This study is based in part on the premise that $\Delta \mathrm{AN}_{12}(f)$ is independent of wind speed. However, it was observed that the variability of $\Delta \mathrm{AN}_{12}$ for the New Jersey and the deep water reference measurements increased as the wind speed decreased. This variability may be partly ascribed to two properties of AN levels. First, as the wind speed decreases the ability to separate wind-generated and shipping-generated noise decreases. This is itself is a source of error in extracting $\Delta \mathrm{AN}_{12}$ from measured data. Second, it is well known that wind-generated AN levels increase rapidly with increasing wind speed at the lower wind speeds but have an asymptotic behavior as the wind speed increases beyond about 30 knots.[6] Thus errors in extracting $\Delta \mathrm{AN}_{12}$ can be mitigated by using AN levels from the high wind speed region. The measured $\Delta \mathrm{AN}_{12}$ values presented in this study are an average over AN levels at wind speeds of 35 and 40 knots.

Figure 3 compares the measured $\Delta \mathrm{AN}_{\mathrm{NJCS,CO}}$ and modeled $\Delta L_{\mathrm{NJCS,CO}}$ values. The $\Delta \mathrm{AN}_{\mathrm{NJCS,CO}}$ values lie approximately between 6 and 8 dB in the 400–3200 Hz band, and the $\Delta L_{\mathrm{NJCS,CO}}$ values also lie between about 6 and 8 dB. In this sense the modeled results are in good agreement with the measured values. Figure 3 includes the result of making the frequency

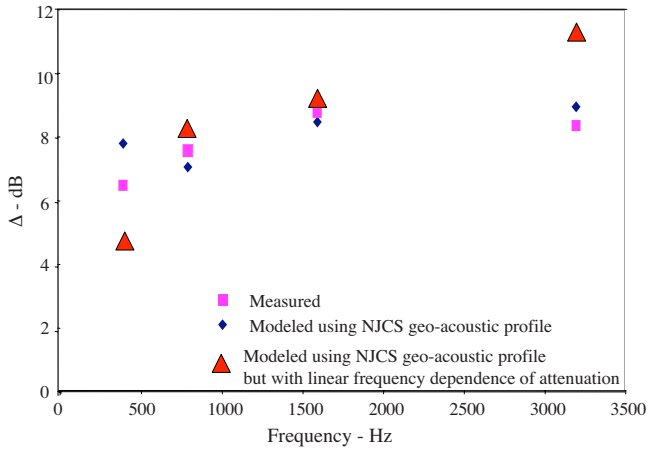


Fig. 3. (Color online) Comparison of $\Delta AN_{NJCS,CO}$ to $\Delta L_{NJCS,CO}$

dependence of the attenuation for the New Jersey site seabed linear with α=0.55 dB/m at 1 kHz. An attenuation that depends linearly with frequency causes the modeled $\Delta L_{NJCS,CO}$ to have a spread of about 6 dB over the 500–3000 Hz band, whereas the spread of measured $\Delta AN_{NJCS,CO}$ values is only about 2 dB apart. Assuming that the sediment sound speed values deduced from the previous analysis[5] are approximately correct, a nonlinear frequency dependence of the sediment attenuation appears to be the reason that the $\Delta AN_{NJCS,CO}$ values have a weak frequency dependence in the 500–3000 Hz band.

5. Summary

Measured ambient noise levels on the New Jersey continental shelf were compared to levels from deep-water reference measurements. The differences were modeled with knowledge of the waveguide properties provided by previous measurements and analyses for the New Jersey shelf site and a specific site in the NE Pacific. The measured AN levels for the New Jersey site measurements are about 6–8 dB above the reference deep water measured levels in the 500–3000 Hz band. Using the properties of the New Jersey and NE Pacific waveguides, the AN level differences were modeled with high-angle propagation approaches. Good agreement was observed between the modeled and measured AN level differences. The calculated results are consistent with a previous analysis of acoustic data that concluded the New Jersey shelf seabed in the proximity of the SWAMI-32 array is composed of a high-speed sandy sediment with a nonlinear frequency dependent attenuation.

Acknowledgments

This work was supported by the U.S. Office of Naval Research, Code 321 OA. The first author wishes to acknowledge helpful discussions with Dr. Ben Reeder.

References and links

[1]V. O. Knudsen, R. S. Alford, and J. W. Emling, "Survey of underwater sound, Report No. 3, Ambient Noise," Office Sci. Res. Develop. Rept. 4333 (1944).
[2]G. M. Wenz, "Acoustic ambient noise in the ocean: Spectra and sources," J. Acoust. Soc. Am. **34**, 1936–1956 (1962).
[3]N. Ross, *Mechanics of Underwater Noise* (Pergamon, Oxford, 1976), pp. 69–71.
[4]F. Ingenito and S. N. Wolf, "Site dependence of wind-dominated ambient noise in shallow water," J. Acoust. Soc. Am. **85**, 141–146 (1989).
[5]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**(3), (2008).
[6]R. D. Gaul, D. P. Knobles, J. Shooter, and A. F. Wittenborn, "Ambient noise analysis of deep ocean measurements in the Northeast Pacific," IEEE J. Ocean. Eng. **32**, 497–512 (2007).
[7]W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J.

Acoust. Soc. Am. **67**, 1988–1996 (1980).
[8]G. B. Morris, "Depth dependence of ambient noise in the Northeastern Pacific Ocean," J. Acoust. Soc. Am. **64**, 581–590 (1978).
[9]A. J. Perrone, "Ambient noise spectrum level as a function of water depth," J. Acoust. Soc. Am. **48**, 362–370 (1970).
[10]E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," J. Acoust. Soc. Am. **100**, 3631–3645 (1996).
[11]E. K. Westwood and R. A. Koch, "Elimination of branch cuts from the normal-mode solution using gradient half spaces," J. Acoust. Soc. Am. **106**, 2513–2523 (1999).
[12]E. K. Westwood and P. J. Vidmar, "Eigenray finding and time series simulation in a layered-bottom ocean," J. Acoust. Soc. Am. **81**, 912–924 (1987).

Field measurements and modeling of attenuation from near-surface bubbles for frequencies 1–20 kHz

Peter H. Dahl
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698
dahl@apl.washington.edu

Jee Woong Choi
Department of Environmental Marine Sciences, Hanyang University, 1271 Sa-3-dong, Ansan, Korea
choijw@hanyang.ac.kr

Neil J. Williams and Hans C. Graber
University of Miami-RSMAS-AMP, 4600 Rickenbacker Causeway, Miami, Florida 33149
nwilliams@rsmas.miami.edu, hans@miami.edu

Abstract: Measurements of excess attenuation from near-surface bubbles from the Shallow Water '06 experiment are reported. These are transmission measurements made over the frequency range 1–20 kHz, and they demonstrate a frequency, grazing angle, and wind speed dependence in attenuation. Data modeling points to bubble void fractions of order 10^{-6} in effect for wind speeds 10–13 m/s. Simultaneous measures of wind speed made within 1.5 and 11 km of the open water experimental location differed by 2 m/s in their respective 30 min average; this has cautionary implications for empirical models for bubble attenuation that are a strong function of wind speed.

PACS numbers: 43.30.Hw, 43.30.Ft [WC]
Date Received: March 13, 2008 **Date Accepted:** June 11, 2008

1. Introduction

Bubbles residing at sufficiently high concentrations near the air–sea interface can impact the performance of acoustic systems in terms of scattering, attenuation, and sound-speed dispersion.[1] For example, for lower grazing angles ($<30°$) and frequencies of O(10) kHz and above, bubbles are the dominant source of backscattering from the air–sea interface[2,3] and their impact is thus readily observable in measurements made even at very low wind speeds (<4 m/s) associated with a mild disruption of the air–sea interface (or microbreaking).[4]

Fewer observations exist showing the effects of attenuation from near-surface bubbles. Estimates that can be linked entirely to the effects of bubbles are generally more difficult to extract from field data, as measurements are often encumbered with a large variance associated with other random features of acoustic transmission loss (TL). Measurements over 10-km (Ref. 5) and 17-km (Ref. 6) paths in shallow water (1–3 kHz in each case) do show an excess attenuation in TL upon the wind speed exceeding ~10 m/s. In the latter case it was postulated that near-surface bubbles were responsible for the excess attenuation, with the attenuation going as a function of frequency, f, and wind speed, U, as $f^{1.5}\,U^4$. However, converting long-range TL estimates into a more generalized metric, e.g., loss in dB/m, requires assumptions on the precise nature of the paths (eigenrays) responsible for signal delivery.

In this paper, which originates from the Shallow Water '06 experiment (SW06), measurements of attenuation associated with bubbles made over a 200-m propagation path are reported over a frequency range 1–20 kHz (frequencies measured simultaneously). The observable is the excess attenuation measured in the surface bounce path (in excess of spreading and seawater absorption), also referred to as surface bubble loss (SBL) and expressed in dB. The

surface bounce path is associated with a very stable eigenray well isolated in time from other arrivals such as the direct path. With, as in this case, details of the transmission path known, a more general metric, β_I, equal to the depth-integrated bubble extinction cross section per unit volume,[2] is readily derived from the SBL estimates, which can be applied to evaluate excess attenuation for other transmission geometries. We examine and contrast 8 h of continual measurements made during SW06 on 10 Aug. (wind speed range 4–7 m/s) with 8 h of measurements made on 15 Aug. (wind speed range 10–13 m/s). During this period continuous measurements of wind, wave, and air–sea stability conditions were made from two nearby moorings.

2. Experimental description

The acoustic observations were made by the University of Washington Applied Physics Laboratory from aboard the research vessel R/V KNORR. An acoustic source (1–20 kHz) was deployed from the stern of the R/V KNORR at depth 40 m, and signals were recorded on a moored receiving system with a remotely changeable configuration. In this study signals received on omni-directional receivers at depths 25 and 50 m are analyzed. The receiving system (referred to in this and other papers on SW06 as MORAY) was deployed at location 39.0245 N, 73.0377 W (depth 80 m), which defined the central (midfrequency) site for SW06 experimental observations, and henceforth is referred to as location M. The R/V KNORR positioned itself at four stations, each at a range 200 m from location M and separated in direction between source and receiver by 90° in bearing. Station-keeping by the R/V KNORR was made precise by its dynamic positioning system, and repeated measurements from these four stations over the course of 7 days showed fluctuations in range of less than 5 m.

An *in situ* acoustic system calibration was carried out at four stations each at range 50 m from the MORAY and separated by 90° in source–receiver bearing angle. The calibration combined both transmitting and receiving system parameters into a single, integrated system parameter (ISP) that depends on frequency and receiver channel (hydrophone). The variance of ISP over the course of 5 days of calibration measurements including the four bearing angles (60 independent measurements at each frequency) forms one component of the experimental error. The transmitted signal was a 3-ms cw pulse, for which all frequencies were superimposed and transmitted simultaneously; a particular frequency is recovered in postprocessing via digital bandpass filtering (acoustic data sampled at 50-kHz sampling rate). The average squared voltage within the time arrival associated with the surface bounce path is taken based on a 20-ping average. This is proportional to a linear intensity average with results converted to dBV, and the estimate of SBL is equal to ISP–dBV–TL, where TL is computed using ray theory and includes the effects of seawater absorption. The conductivity-temperature-depth (CTD) measurements from the R/V KNORR were used to compute the ray theory estimates of TL and sea surface grazing angle, θ_g. As a check on the stability of the eigenray paths due to the changing sound-speed profile, a ray analysis was done for each day from 10–15 Aug. using 15-min-averaged sound-speed profiles derived from temperature measurements taken every 30 s by a nearby mooring from the Woods Hole Oceanographic Institution. The result is that TL varied by <0.5 dB and θ_g by $<0.5°$.

The primary acoustic measurement is illustrated in Fig. 1, showing typical sound-speed profiles [Fig. 1(a)], and the corresponding ray diagram [Fig. 1(b)] for the first five eigenrays delivering the signal over the 200-m transmission range. The sea surface bounce path (*S*) is the subject of this study, with other waterborne arrivals identified in the caption. The time series [Fig. 1(c)] represents the average of a 20-ping transmission set (center frequency 20 kHz) made on 10 Aug. 12:30 UTC (solid line) and 15 Aug. 16:50 UTC (dashed line). The time arrival structure for the sea surface bounce path is shaded, and the calibrated excess attenuation estimates (SBL) for 10 Aug. and 15 Aug. are 0 and 7 dB, respectively. To more effectively illustrate the significance of attenuation in this display, the 10 Aug. received level (identified by symbol) is mapped to 0 dB and the 15 Aug. time series is given the same offset.

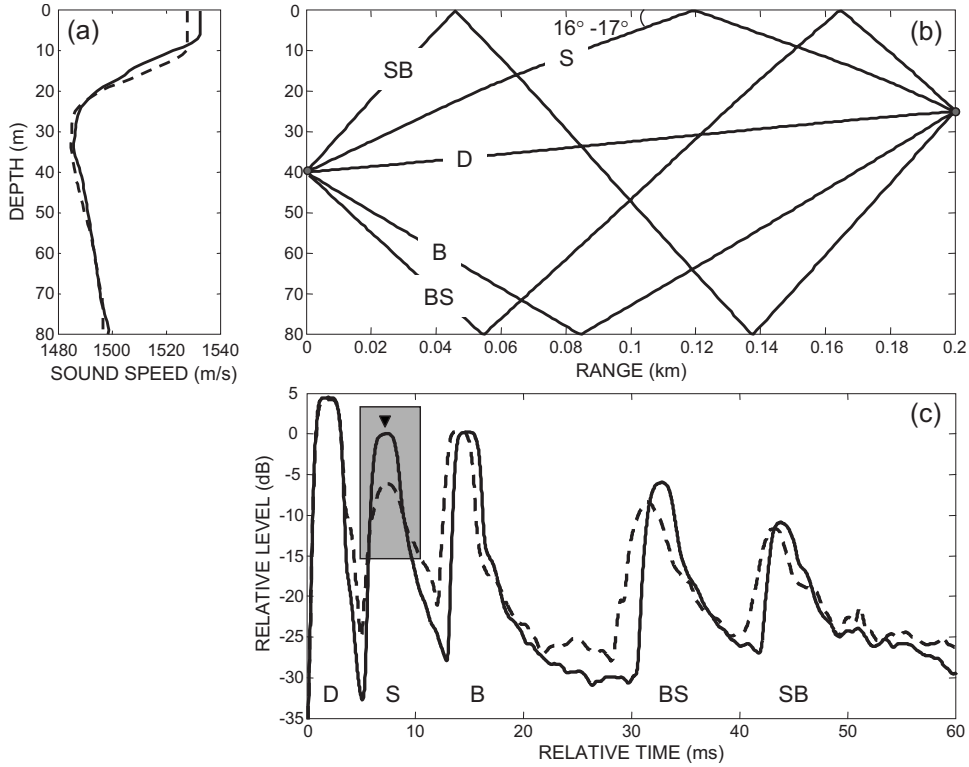


Fig. 1. (a) Representative sound-speed profile for 10 Aug. (solid line) and 15 Aug. (dashed line) (b) Ray diagram for source at 40 m, 25-m receiver depth, and range 200 m, showing the first five eigenrays computed using the 15 Aug. sound-speed profile (10 Aug. profile yielding similar result). Waterborne paths in order of their arrival are the direct (D), surface (S), bottom (B), bottom-surface (BS), and surface-bottom (SB). The second arriving sea surface bounce path (*S*) is the subject of this study and the sea surface grazing angle, θ_g, range is noted (it is the third arrival for 50-m receiver depth, with $\theta_g = 22-23°$). (c) Time series of relative received level for a pulse with 20-kHz center frequency, based on the average of 20-ping transmissions on 10 Aug. 12:30 UTC (solid line) and 15 Aug. 16:50 UTC (dashed line), with waterborne paths identified as in (b). The time arrival structure for sea surface bounce path is shaded, and the calibrated excess attenuation estimates for 10 Aug. and 15 Aug. were 0 and 7 dB, respectively. Arrival time is relative to that of the direct path (135 ms). For this display, the 10 Aug. received level (at symbol) is mapped to 0 dB and the 15 Aug. time series is given the same offset.

Air–sea measurements are derived from the two (Yankee and Romeo) Air–Sea Interaction Spar (ASIS) buoy[7] moorings from the University of Miami. The Yankee mooring was 1.5 km from the acoustic receiving array at location M, and the Romeo mooring was 11 km shoreward. In view of the open water conditions, both mooring-based wind speed measures, based on a 30-min average, provided an equally representative measure of wind speed. (A reasonable conjecture is that turbulence/eddies in the atmosphere travel with the mean wind speed, and for wind speed in the range 10 m/s the two ASIS buoys require temporal averaging times of order 10^3 s for comparison.) Similar 30-min averages are taken of measurements of air and sea temperature, relative humidity, and air pressure from the ASIS buoys, and these are used to convert the wind measurements taken at approximately 6 m above sea level to the equivalent neutral winds at 10 m height (U_{10n}) by using an assumed logarithmic wind profile.[8]

3. Results and modeling

Figure 2 summarizes the results for 1, 4, 12, 16, and 20 kHz for the 10 Aug. measurements (left) and 15 Aug. measurements (right). Wind speed measured at the Yankee (black) and Romeo (gray) ASIS moorings, compared with attenuation (SBL) in dB, are shown referenced to the left

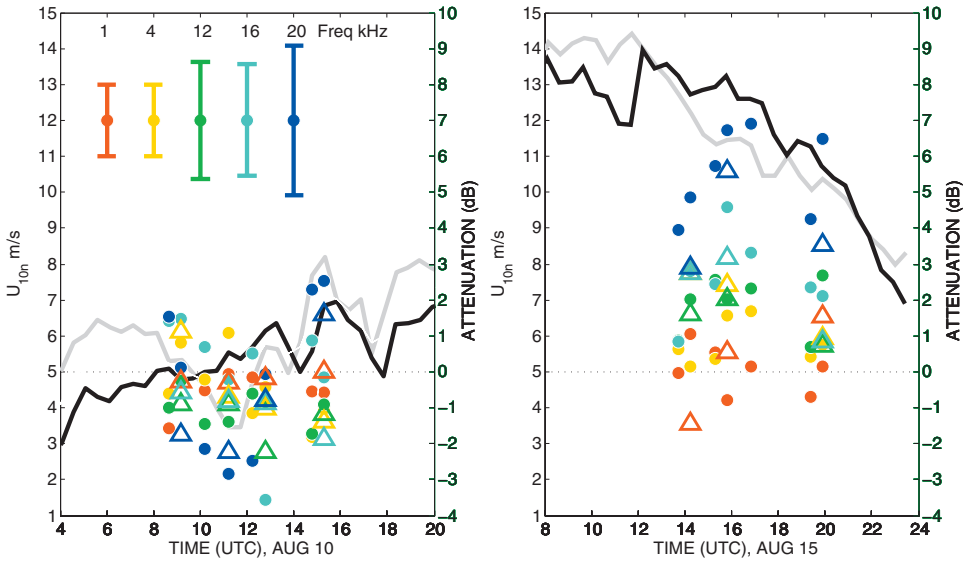


Fig. 2. Summary of attenuation measurements at 1, 4, 12, 16, and 20 kHz, for 10 Aug. (left) and 15 Aug. (right), plotted against the Yankee (black) and Romeo (gray) wind speed measurements from the ASIS moorings. Ten August measurements correspond to grazing angle 16° (closed circle) and 22° (open triangle); 15 Aug. measurements follow same convention with grazing angles increased by ~1°. Representative frequency-dependent error bars are shown in the left plot and color code applies to both grazing angles and both days. Note that each plot has dual vertical axes and the dotted horizontal line identifies a 0-dB reference but otherwise has no relation to wind speed. For reference, sunrise and sunset are 09:50 and 23:50, respectively.

and right vertical axis, respectively (the horizontal dotted line provides a reference to 0-dB attenuation). The 10 Aug. measurements correspond to grazing angle 16° (closed circle) and 22° (open triangle), and 15 Aug. measurements follow the same convention with grazing angles increased by ~1° owing to changes in sound speed caused by bubbles (discussed below). Representative frequency-dependent error bars are shown in the left plot and apply to both days.

The 10 Aug. attenuation measurements are nominally encompassed by a 2-dB spread about 0 dB, with a few measurements exceeding this range; 0 dB is the expected value assuming there are no additional sources of attenuation beyond the ray-based TL estimate, plus a small (frequency-dependent) contribution due to absorption in seawater. The rms wave height H is ~0.16 m over the course of the 10 Aug. measurements, and the roughness parameter $\chi = 2kH \sin \theta_g$ exceeds 1.5 at 4 kHz and therefore incoherent intensity dominates the arrival from the sea surface bounce path for frequencies >4 kHz. For frequencies of O(10) kHz, χ is $\gg 1$ and any coherent component in intensity is now vanishingly small. The incoherent intensity will tend towards an exponential distribution in probability; a variable representing the linear intensity average of $N=20$ realizations expressed in decibels will exhibit characteristics of a Gaussian probability density function (PDF) with standard deviation ~1 dB. The aforementioned ISP sample standard deviation is frequency dependent, ranges from 0.5 dB at 1 kHz to 2.2 dB at 20 kHz, with measurements characterized by a symmetric empirical PDF. The variances from these two independent errors add, and thus a ±2-dB spread is expected along with the observed trend of increasing spread with frequency, as reflected in the error bars.

For the 15 Aug. attenuation measurements, statistically significant observations are made that tend to increase with frequency. Note that H has increased to ~0.4 m with χ exceeding 1.5 for all frequencies; however, we assume this parameter to have little influence given that its frequency dependence diminishes once the parameter exceeds unity. Estimates of attenuation made on the upper, 25-m-depth receiver (lower grazing angle) are in general higher than

those made simultaneously on the lower, 50-m-depth receiver (higher grazing angle). This property and the observed frequency dependence can be understood with modeling.

To model these results we take the bubble-mediated transmission loss, TL_{bub}, for signals propagating through a path traversing a bubbly region to be of the form,[3]

$$TL_{bub} = 4.34 \int_{path} s_e(f,z)dl, \tag{1}$$

where s_e is the near-surface bubble population's extinction cross section per unit volume,[1] as determined by the integral over the product of the bubble size distribution, $N(a)$, and the extinction cross section for a single bubble of radius a. We postulate a functional form for $N(a)$ defined as follows:

$$n(a) = 1 \quad \text{for } a < 20\ \mu\text{m}, \tag{2a}$$

$$n(a) = (a/20\ \mu\text{m})^{-5.5}, \quad 20\ \mu\text{m} < a < 800\ \mu\text{m}, \tag{2b}$$

$$n(a) = n(800\ \mu\text{m})\ (a/800\ \mu\text{m})^{-7}, \quad a > 800\ \mu\text{m}, \tag{2c}$$

where $N(a)$ differs from $n(a)$ by a multiplicative constant determined by the void fraction. The size distribution for $a < 20\ \mu$m is not critical given the frequency range. We assume that bubbles associated with a more stabilized near-surface layer are responsible for the attenuation, rather than nascent bubbles associated with an actively breaking wave; and a power law used for bubbles with radii greater than 20 μm is typical for oceanic bubbles in this age class.[9] For larger bubbles, of order 1 mm radius, we recognize a Hinze scale,[10] and place a break point in the bubble size distribution here. A maximum radius of 1600 μm is used.

The path in Eq. (1) is the portion of the surface bounce path that traversed the near-surface layer of bubbles. Assuming a depth dependence $s_e(z) = s_e(0)\exp(-z/L)$, where L is an e-folding depth scale, means integration along the path equals $2\ \beta_I/\sin(\theta_g)$, where $\beta_I = Ls_e(0)$. The measurements of SBL at a given grazing angle θ_g can thus be converted directly to β_I, and upon assuming a depth scale, to $s_e(0)$. The depth scale L is typically O(1) m, and we use $L = 1.4$ m arising from the large set of observations in Ref. 3.

Figure 3 shows attenuation measurements for 10 Aug. (left-side plots) and 15 Aug. (right-side plots) averaged into two wind speed classes for each day: defined as 5 and 7 m/s for 10 Aug. and, 11 and 13 m/s for 15 Aug., where a class includes all acoustic measurements made at wind speeds within ± 1 m/s of its defined value. (An average of the Yankee and Romeo wind measures is used for this purpose.) The model for the 15 Aug. data requires a bubble void fraction of 6.5×10^{-7} and the model (black, solid line) tends to capture both level and frequency dependence in the data. The dashed lines are the model with factor of 2 changes in void fraction. One notable difference between model and data is the >2-dB attenuation at 4 kHz for the 50-m receiver depth. It is possible that bio-absorption due to swim bladder-bearing fish[11] may be in effect, although 4-kHz measurements made at the same time (16:00 UTC) on 10 Aug. do not exceed 2 dB. Also, with the order 10^{-6} void fraction in effect for 15 Aug., there is a bubble-mediated sound-speed change.[1] Using the above bubble size distribution, there is a sound-speed reduction of $\sim$10 m/s within L-m of the air–sea interface; the ensuing upward refraction changes the grazing angles slightly as noted in Fig. 3.

4. Conclusions

The SW06 data reported here demonstrate a frequency, grazing angle, and wind speed dependence in excess attenuation from near-surface bubbles. These represent the first transmission-type measurements of bubble attenuation made over a broad frequency range (1–20 kHz, made simultaneously), for which the signal delivery path is known. The *total acoustic field* for a single sea surface bounce path is studied by taking the average of the squared envelope of pressure (voltage). Five conclusions emerge: (1) a statistically significant attenuation is not observable

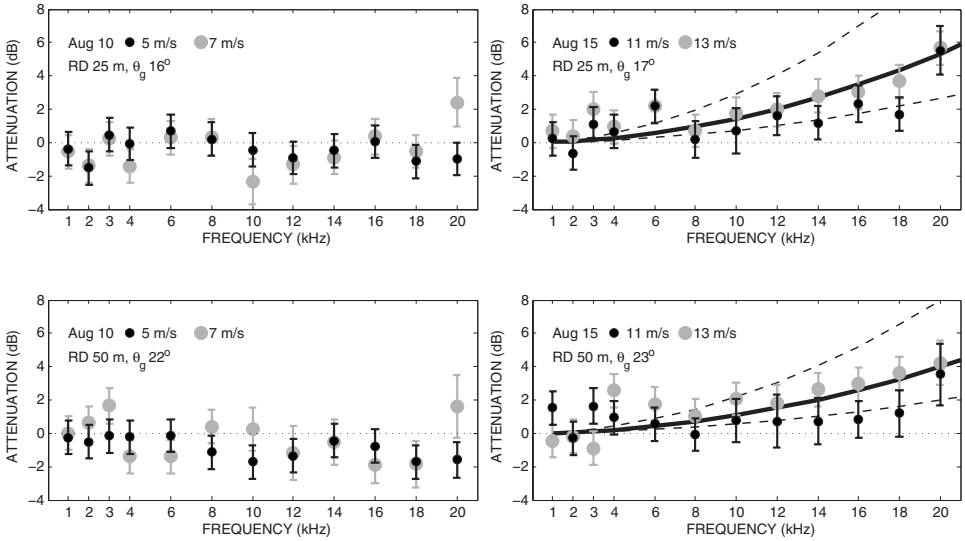


Fig. 3. Summary of attenuation measurements made on 10 Aug. (left plots) and 15 Aug. (right plots) as averaged into the wind speed classes defined in the text. Model for measurements on 15 Aug. (thick, black line) uses a void fraction of 6.5×10^{-7}, with dashed lines representing the same model but with void fraction varying by a factor of 2.

until wind speeds exceed the 8–10-m/s range. This is consistent with previous studies made with a smaller frequency range. (2) For higher wind speeds a frequency-dependent attenuation goes approximately as $f^{2\text{ to }2.5}$, where f is frequency. This reflects an underlying bubble size distribution that goes as $a^{-5\text{ to }-5.5}$, where a is bubble radius, and is consistent with modeling assumptions. (3) Attenuation measurements at $\theta_g=23°$ were lower than at $\theta_g=17°$, which supports the postulation that, in a time-averaged sense, attenuation arises from a thin, near-surface layer of bubbles of depth scale $L=\mathrm{O}(1)$ m. (4) The ability to specify oceanic bubble void fraction is arguably limited to its magnitude, yet changes in attenuation scale directly with void fraction as shown by the results of the factor of 2 changes in void fraction. Finally (5) the two ASIS wind speed measures, equally valid in their representation of average open-water air–sea conditions, differed by ~2 m/s over their respective 30-min averages. Given that empirical model functions for bubble attenuation typically involve wind speed in a power law or similarly strong function, a high degree of model prediction uncertainty is to be expected.

Acknowledgments

This research was supported by the Office of Naval Research. The first author acknowledges the data from the temperature mooring used for simulation, made available courtesy of the Woods Hole Oceanographic Institution, Ocean Acoustics Laboratory.

References and links

[1]C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications*, Wiley-Interscience, New York (1977).
[2]P. H. Dahl, W. J. Plant, B. Nützel, A. Schmidt, H. Herwig, and E. A. Terray, "Simultaneous acoustic and microwave backscattering from the sea surface," J. Acoust. Soc. Am. **101**(5), 2583–2595 (1997).
[3]M. V. Trevorrow, "Measurements of near-surface bubble plumes in the open ocean with implications for high-frequency sonar performance," J. Acoust. Soc. Am. **114**, 2672–2684 (2003).
[4]P. H. Dahl, "The contribution of bubbles to high-frequency sea surface backscatter: A 24-h time series of field measurements," J. Acoust. Soc. Am. **113**, 769–780 (2003).
[5]P. C. Wille and D. Geyer, "Simultaneous Measurements of Surface Generated Noise and Attenuation at the Fixed Shallow Water Range NORDSEE," in Proceedings Advanced Research Workshop on Natural Mechanisms of Surface Generated Noise in the Ocean, June 1987, Lerici, Italy, pp. 295–308 (1987).
[6]D. E. Weston and P. A. Ching, "Wind effects in shallow-water transmission," J. Acoust. Soc. Am. **86**, 1530–1545 (1989).
[7]H. C. Graber, E. A. Terray, M. A. Donelan, W. M. Drennan, J. C. Vanleer, and D. B. Peters, "ASIS—A new air-

sea interaction spar buoy: Design and performance at sea," J. Atmos. Ocean. Technol. **17**, 708–720 (2000).
[8]W. M. Drennan and L. K. Shay, "On the variability of the fluxes of momentum and sensible heat," Boundary-Layer Meteorol. **119**, 81–107 (2006).
[9]D. M. Farmer and S. Vagle, "Waveguide propagation of ambient sound in the ocean-surface bubble layer," J. Acoust. Soc. Am. **86**, 1897–1908 (1989).
[10]G. B. Deane and M. D. Stokes, "Scale dependence of bubble creation mechanisms in breaking waves," Nature (London) **418**, 839–844 (2002).
[11]O. Diachok, B. Liorzou, and C. Scalabrin, "Estimation of the number density of fish from resonance absorptivity and echo sounder data," ICES J. Mar. Sci. **58**, 137–153 (2001).

Shallow-water seismoacoustic noise generated by tropical storms Ernesto and Florence

James Traer, Peter Gerstoft, Peter D. Bromirski, William S. Hodgkiss, and Laura A. Brooks[a)]
Scripps Institution of Oceanography, La Jolla, California 92093-0238
jtraer@ucsd.edu, gerstoft@scsd.edu, peter@coast.ucsd.edu, whodgkiss@ucsd.edu, lbrook02@gmail.com

Abstract: Land-based seismic observations of double frequency (DF) microseisms generated during tropical storms Ernesto and Florence are dominated by signals in the 0.15–0.5 Hz band. In contrast, data from sea floor hydrophones in shallow water (70 m depth, 130 km off the New Jersey coast) show dominant signals in the ocean gravity-wave frequency band, 0.02–0.18 Hz, and low amplitudes from 0.18 to 0.3 Hz, suggesting significant opposing wave components necessary for DF microseism generation were negligible at the site. Florence produced large waves over deep water while Ernesto only generated waves in coastal regions, yet both storms produced similar spectra. This suggests near-coastal shallow water as the dominant region for observed microseism generation.

PACS numbers: 43.30.Nb, 43.30.Pc, 91.30.Ye [WC]
Date Received: April 1, 2008 **Date Accepted:** June 19, 2008

1. Introduction

Microseisms are ubiquitous seismic signals generated by ocean waves.[1] The peak of the microseism spectrum occurs near twice that of ocean surface waves [double frequency (DF) microseisms], generated by the interaction of opposing surface waves of nearly the same wave number.[2] Unlike traveling ocean waves which decay exponentially with depth, the amplitude of the DF pressure pulse does not decay appreciably with depth.[2] Primary microseisms are observed at ocean wave frequencies and are generated only in shallow water by breaking waves or interaction with the sloping bottom.[3] The results from beamforming with seismic arrays suggests that the dominant source regions for primary and double frequency signals may differ in space and/or time.[4]

Recent work suggests that microseisms are only generated when surface waves approach coastal areas[5–8] and that the generation of microseisms is well correlated with ocean surface conditions.[5,9] Storms over the ocean generate large waves causing elevated microseism levels that have been attributed to specific storm events.[4,10]

The surface gravity-wave induced pressure at the sea floor is[5]

$$p = \frac{p_0}{\cosh kH}, \tag{1}$$

where p_0 is the pressure at the surface, H-ocean depth, and k the wave number, determined by the surface gravity wave dispersion, $\omega^2 = gk \tanh kH$, where ω is the angular frequency, $\omega = 2\pi f$, with f the surface wave frequency. For example, with $H = 70$ m (as in the experiment discussed), the pressure from a 0.1 Hz wave will be $p = 0.1p_0$ and the pressure from a 0.3 Hz wave will be $p = 10^{-11}p_0$. This indicates that the decay of wave pressure with depth is strongly dependent on the wave frequency, and that the pressure spectrum from overhead wave activity observed at the bottom in a shallow-water environment will be dominated by direct forcing

[a)]Also at the School of Mechanical Engineering, University of Adelaide, Adelaide, Australia.

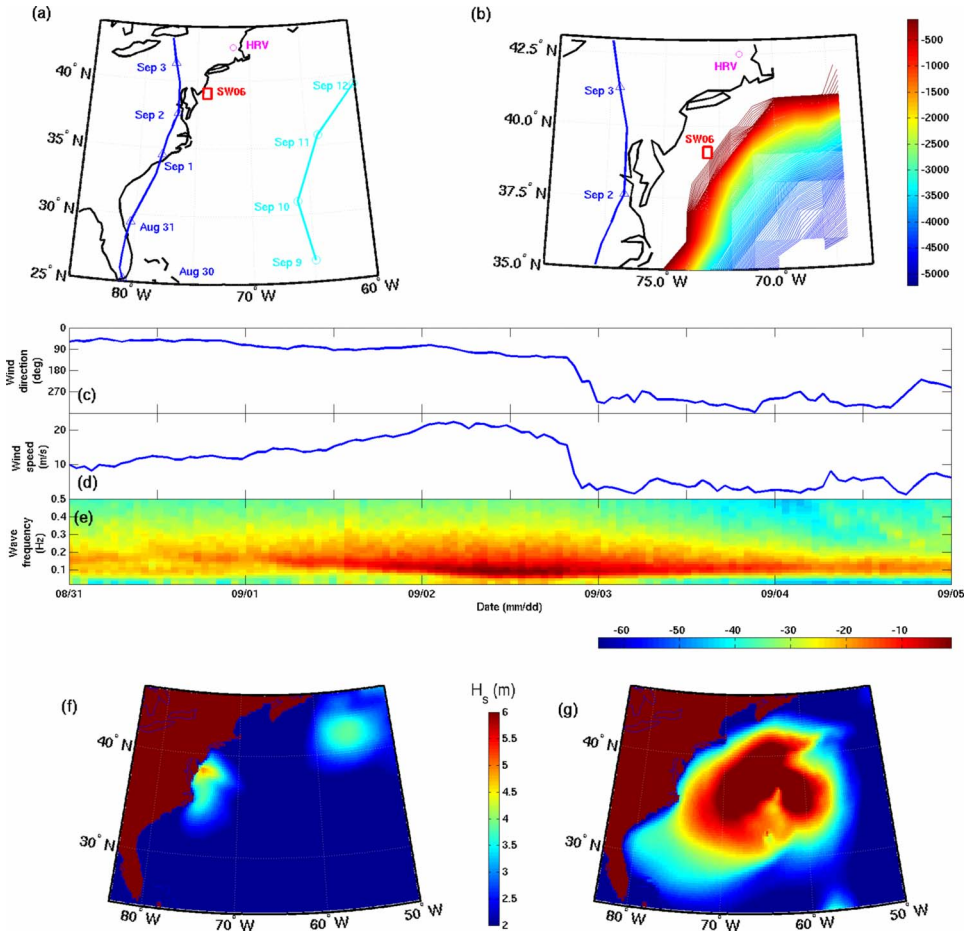


Fig. 1. (Color online) The experiment environment. (a) Experiment location (rectangles) and the recorded path of the storm centers. Triangles mark the storm center for Ernesto and circles the center for Florence every 24 h starting 0 Z 30 August and 9 September, respectively. (b) Bathymetry contours from 100 to 5000 m depth. Water depth less than 100 m is white. (c) Wind direction. (d) Wind speed. (e) The surface wave spectra (dB) from 0.02 to 0.5 Hz for 30 August–3 September. The wave spectral energy is normalized with respect to the highest observed signal. Wind and wave data from the ASIS buoys are averaged over $\frac{1}{2}$ h periods. (f), (g) Significant wave heights (H_s) from Ernesto (9 Z 2 September) and Florence (9 Z 12 September), respectively.

from low frequencies. DF pressure fluctuations will become dominant in deep water because they do not decay appreciably with depth regardless of their frequency.

An opportunity to study storm-generated microseisms on both the ocean bottom and land occurred when waves generated by Tropical Storms Ernesto and Florence passed over the Shallow Water Experiment 2006 (SW06) site during September 2006. Sea floor hydrophones from the SWAMI32, SWAMI52, and SHARK arrays measured pressure variations at the ocean bottom on the shallow-water continental shelf. These were compared with broadband seismic data from the HRV (Harvard) station in Massachusetts.

2. Array environment

The edge of the leading right-hand quadrant of Tropical Storm Ernesto passed over the arrays on 2 September, while the inland storm center moved northward [Fig. 1(a)]. The storm, recorded by University of Miami Air-Sea Interaction Spar (ASIS) Buoys,[11] generated steady high winds and wave energy [see Figs. 1(c)–1(e)] over the arrays for 2 days. The wind speed sud-

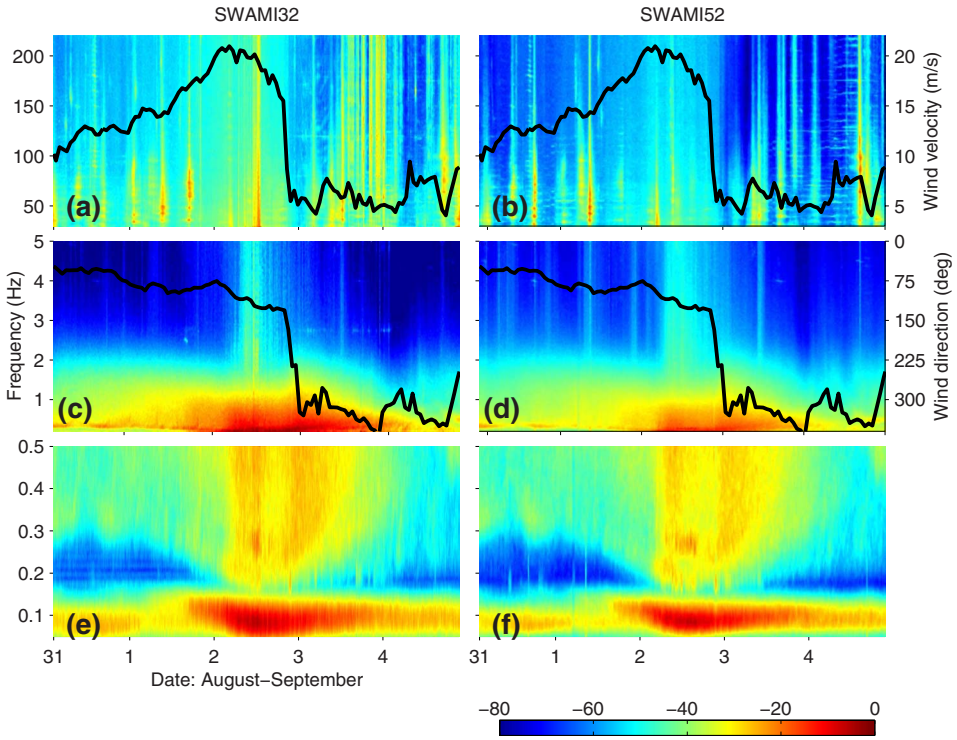


Fig. 2. (Color online) Normalized spectrograms (dB) of the acoustic data at three frequency scales [(a), (b) 30–220 Hz, (c), (d) 0.2–5 Hz, and (e), (f) 0.02–0.5 Hz] obtained over a 5 day period (31 August–4 September) from the SWAMI32 [(a), (c), and (e)] and SWAMI52 [(b), (d), and (f)] arrays. The wind velocity trace from Fig. 1(d) is superimposed in (a) and (b) and the wind direction from Fig. 1(c) is superimposed in (c) and (d). The spectrograms are averaged over five hydrophones and normalized with respect to the highest power spectral density in the observed range.

denly dropped and changed direction near the end of 2 September as the storm passed the arrays [Figs. 1(c) and 1(d)]. Throughout 1 and 2 September the winds blew from the east with speeds from 10 to 20 m/s. From 20 Zulu (20 Z) 2 to 3 September the winds blew from the west at speeds from 2 to 10 m/s. Wave energy remained strong for a further 28 h after the wind dropped [Fig. 1(e)]. The waves underwent a change in direction over 2 September (not shown), transitioning from eastward to south-west incidence, potentially producing an opposing wave field for standing wave generation and transmission of microseisms to the seabed.[2] NOAA hindcasts[12] showed that Ernesto produced large waves in shallow water such as the SW06 site, but not in deep waters [Fig. 1(f)]. The large waves in Fig. 1(f) at (42°N, 55°W) were from another storm. Wave data from the ASIS buoy were consistent with the hindcast results at the site.

On 10 September, large waves from Tropical Storm Florence arrived at the site. Florence moved northward through the Atlantic Ocean with the storm center remaining in deep water [Figs. 1(a) and 1(g)]. The ASIS buoys and both SWAMI arrays were removed prior to the arrival of Florence and no wind or wave data were available. The SHARK array recorded acoustic data through 14 September. The SWAMI and SHARK arrays were situated on a sandy floor at a depth of 70–80 m, 20 km west of the continental shelf and 130 km from the New Jersey coast [Fig. 1(b)].

Although the hydrophones were not designed to work at frequencies less than 2 Hz, the relative spectrogram levels (Fig. 2) were corrected using the SWAMI hydrophone frequency response and the characteristics of the electronic filters. From 10 to 15 Z, 2 September, the

SWAMI32 hydrophones recorded broadband clicks likely attributed to motion of the array. This distortion was minimized by excising any segments with an amplitude greater than 4 s.d. of the signal within each 6 min, 24 sec file. For the SHARK array any segment with an amplitude greater than 8 s.d. was excised. The frequency response of the SHARK hydrophones is unknown. The SHARK frequency response was estimated by comparing the spectra obtained by the SHARK and SWAMI52 arrays over the period of time when the two arrays overlapped (18 Z 25 August to 17 Z 6 September) and assuming the two arrays are measuring the same signal.

3. Acoustic spectrograms

At frequencies above 30 Hz, the acoustic levels are well correlated with the local wind speed, while below 2 Hz a significant signal is observed for 2 days after the passing of the storm [Figs. 2(a)–2(d)]. This suggests the signals observed above 30 Hz are generated by wind-induced breaking waves, in agreement with others,[13] and signals below 2 Hz are generated by surface waves. Wave–wave interactions can produce signals as high as 7 Hz,[14] and may be the source of the large signals on 2 September from 2 to 5 Hz.

The signals recorded by the two arrays are similar at frequencies below 2 Hz, suggesting that either these signals travel over the 23 km distance separating the arrays, or that the surface wave spectra are similar over many kilometers. The dominant signal occurred at 0.02–0.18 Hz [Figs. 2(e) and 2(f)] throughout the 5-day period, corresponding to the surface wave frequency band [Fig. 1(e)]. This 0.02–0.18 Hz signal is at a maximum when the storm is above the array. Seismic arrays in California detected a strong 0.07–0.11 Hz signal at this time, originating along an azimuth consistent with the signal being generated in coastal waters between 38 and 40°N.[4] This region includes the SWAMI and SHARK arrays. The peak in seismic signal at this time is not correlated with an increase in wave amplitudes, suggesting that microseism generation may be dependent on the location of the waves. Near these times, NOAA hindcasts show large waves impacting the coast of Cape Cod,[12] which has previously been identified as a site of strong microseism generation.[10]

Relatively low spectral levels were observed in the 0.14–0.18 Hz band during times when the peak in wave energy near 0.08 Hz occurred. The frequency range of the low energy band should include high amplitude DF microseism signals. The ASIS buoy measured the dominant wave direction as westward over 1 September and northward over 2 September, toward the coasts of New Jersey and Cape Cod, respectively. Conceptually, opposing waves from coastal reflection interacting with incoming swells should produce standing waves, however, no DFs were observed at magnitudes equivalent to the primary pressure wave. The absence of DF signal associated with the dominant wave frequency indicates that little opposing wave energy was present. Note that DF microseisms on land and in the deep ocean typically have much higher amplitudes than primary microseisms.[5,10,14] DF levels at HRV are much higher than associated primaries during Ernesto, but the opposite is true at the sea floor arrays, indicating that the signal observed by the arrays is dominated by direct pressure from overhead waves, not from microseisms.

The 0.2–2 Hz signal is 20–30 dB weaker than the 0.02–0.18 Hz signal. It does not appear in the wave spectra, which suggests that this signal is likely either DF microseisms, or due to an inaccurate response of the wave buoy at higher frequencies. The signal maximum in this 0.2–2 Hz band occurs at 03 Z 2 September [Figs. 2(c)–2(f)]. A second signal maxima is observed between 0.2 and 2 Hz at 18 Z 2 September and corresponds to a shift in local wind direction [Figs. 2(c) and 2(d)], suggesting that opposing seas forcing necessary for the DF mechanism occurred.

The SHARK array was used to compute spectrograms from 25 August to 14 September [Fig. 3(b)]. As with the SWAMI measurements of Ernesto, during Florence the dominant signal occurred between 0.02 and 0.18 Hz, corresponding to the dominant wave frequency band, with little signal observed in the associated DF band. As Florence did not make landfall, it had a much larger fetch than Ernesto and produced lower frequency ocean waves for a longer time. The lack of DF signals during both tropical storms suggests that opposing wave compo-

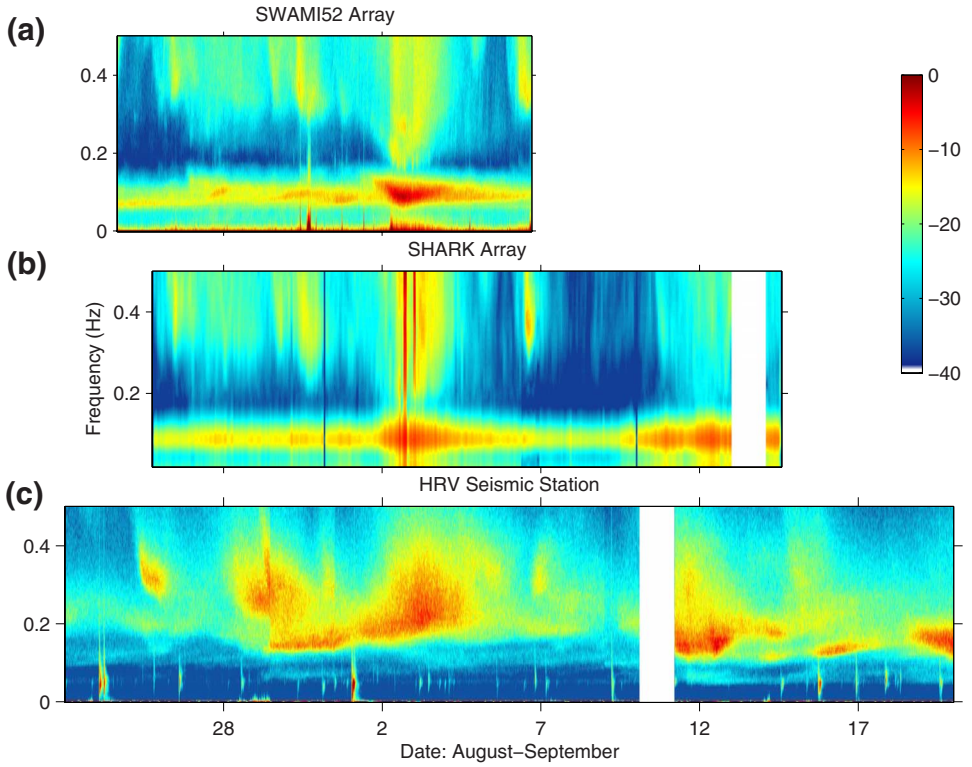


Fig. 3. (Color online) Normalized spectrograms (dB) of the acoustic data between 0.01 and 0.5 Hz from the (a) SWAMI52 array, (b) SHARK array, and (c) HRV seismic station. SWAMI52 data were available from 24 August to 6 September, SHARK data from 25 August to 19 September, and HRV data from 23 August to 20 September.

nents at swell frequencies are not generated by wave reflection/scattering at the coastline nearest the arrays and/or because these components are dissipated by bottom interaction when crossing the relatively wide shallow shelf.

4. Seismic spectrograms

In contrast to the spectrograms from the acoustic arrays, the land-based seismic spectra at HRV are dominated by DF signals between 0.15 and 0.5 Hz, which are 20 dB stronger than the primary signal [Fig. 3(c)]. The microseisms from Ernesto's waves are initially seen at 0.15 Hz on 29 August as Ernesto impacts Florida. The signals are present for 6 days as Ernesto travels northward, increasing in frequency as the storm center moves on land and the fetch of the storm decreases, generating more short period wave energy near shore. The 0.2–0.4 Hz signal present from 29 to 30 August is likely due to a smaller storm to the northeast [Fig. 1(f)]. During the period when Ernesto was over the arrays (2 September), the 0.02–0.18 Hz band that dominated the SWAMI array data was largely absent at HRV, confirming that the signals in that band at the arrays were dominated by direct pressure from overhead waves.

After Ernesto dissipated, another strong set of signals appear from 12 to 14 September, attributed to waves generated by Tropical Storm Florence. The signal from Florence has higher amplitude than that of Ernesto and lower frequency, consistent with the larger waves and fetch of Florence [Figs. 1(f) and 1(g)]. These waves impacted the coasts from Florida to Nova Scotia. As with the seismic signal from Ernesto, the DF signal at HRV from Florence is 20 dB stronger than the primary signal.

5. Discussion

The seismic signals of both storms recorded on land were dominated by the DF signals, which is consistent with previous studies and implies that both storms generate DF waves. However, the acoustic levels of Ernesto and Florence contained a weak signal in the lower part of the DF range (0.18–0.25 Hz). Rather, they are dominated by signals at the same frequency as the ocean waves, consistent with a linear forcing mechanism. Despite both storms producing westward traveling waves, no evidence of a wave–wave interaction due to reflection of waves from the coast was observed. This suggests that low frequency opposing waves and corresponding DF microseisms are negligible near the SW06 site, consistent with observations at HRV during the 1991 "Perfect Storm"[10] and that the DF signals detected by the HRV seismic station were generated elsewhere.

When Ernesto reached the SW06 site, the storm center was inland [Fig. 1(f)] and most of the storm-generated waves were in shallow waters. Ernesto produced a significant DF signal at nearby seismic stations over a much longer period of time than the larger storm Florence, which produced larger waves over a greater area in the deep ocean [Fig. 1(g)]. Babcock *et al.*[9] observed high amplitude DF microseisms with an ocean bottom array at a nearby deep water site, indicating that even smaller storms can generate appreciable DF energy and that waves from Florence and Ernesto were likely generating DF microseisms offshore in the open ocean. Despite these large disparities in the size of storms Florence and Ernesto [Figs. 1(f) and 1(g)] the associated microseism signals were similar [Fig. 3(c)], suggesting only near-shore waves produce the signals detected at the HRV seismic station.

The absence of DF signals at the SW06 site may be due to variations in local bathymetry. The SW06 site is separated from the nearest coast by 130 km of shallow water. Waves below 0.1 Hz interact with the bottom at this depth and may attenuate before reaching the coast. It may be the case that the DF microseisms observed at the HRV seismic station are generated in waters shallower than the SW06 site (70 m). The continental shelf narrows substantially near Cape Hatteras to the south and Cape Cod to the north. These regions may be sites of high microseism generation as low frequency waves may attenuate less and hit the coast with more energy.

Acknowledgments

This project was funded by: the Office of Naval Research, the Department of Energy National Energy Technology Laboratory via the Gulf of Mexico Hydrates Research Consortium, University of Mississippi, and the California Department of Boating and Waterways (to P.D.B.). Data assistance provided by Neil Williams and Hans Graber, University of Miami (local wind and wave data); Arthur Newhall, Woods Hole Oceanographic Institute (SHARK); and Applied Research Laboratories, University of Texas (SWAMI) are appreciated.

References and links

[1]S. C. Webb, "The Earth's 'hum' is driven by ocean waves over the continental shelves," Nature (London) **445**, 754–756 (2006).
[2]M. S. Longuet-Higgins, "A theory of the origin of microseisms," Philos. Trans. R. Soc. London **293**, 1–35 (1950).
[3]K. Hasselmann, "A statistical analysis of the generation of microseisms," Rev. Geophys. **1**, 177–210 (1963).
[4]P. Gerstoft and T. Tanimoto, "A year of microseisms in southern California," Geophys. Res. Lett. **34**, L20304 (2007).
[5]P. D. Bromirski and F. K. Duennebier, "The near-coastal microseism spectrum: Spatial and temporal wave climate relationships," J. Geophys. Res. **107**, 2166 (2002).
[6]A. Friedrich, F. Krüger, and K. Klinge, "Ocean-generated microseismic noise located with the Gräfenberg array," J. Seismol. **2**, 47–64 (1998).
[7]T. Tanimoto, "Excitation of microseisms," Geophys. Res. Lett. **34**, L05308 (2007).
[8]P. Gerstoft, M. C. Fehler, and K. G. Sabra, "When Katrina hit California," Geophys. Res. Lett. **33**, L17308 (2006).
[9]J. M. Babcock, B. A. Kirkendall, and J. A. Orcutt, "Relationships between ocean bottom noise and the environment," Bull. Seismol. Soc. Am. **84**, 1991–2007 (1994).
[10]P. D. Bromirski, "Vibrations from the perfect storm," Geochem., Geophys., Geosyst. **2**, 000119 (2001).
[11]H. C. Graber, E. A. Terray, M. A. Donelan, W. M. Drennan, J. C. Van Leer, and D. B. Peters, "ASIS—A new

air–sea interaction spar buoy: Design and performance at sea," J. Atmos. Ocean. Technol. **17**, 708–720 (2000).
[12]"NOAA wavewatch III," URL http://polar.ncep.noaa.gov/waves/nww3_hist.html, National Oceanic and Atmospheric Administration, downloaded 20 March 2008.
[13]D. P. Knobles, S. Joshi, and R. D. Gaul, "Analysis of wind-driven ambient noise in a shallow water environment with a sandy seabed," J. Acoust. Soc. Am. **124**, EL157–EL162 (1993).
[14]C. S. McCreery, F. K. Duennebier, and G. H. Sutton, "Correlation of deep ocean noise (0.4–30 Hz) with wind, and the Holu Spectrum—A worldwide constant," J. Acoust. Soc. Am. **93**, 2639–2648 (1993).

Modal processing for acoustic communications in shallow water experiment

Andrey K. Morozov, James C. Preisig, and Joseph Papp
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
amorozov@whoi.edu, jpreisig@whoi.edu, jpapp@whoi.edu

Abstract: Acoustical array data from the Shallow Water Acoustics experiment was processed to show the feasibility of broadband mode decomposition as a preprocessing method to reduce the effective channel delay spread and concentrate received signal energy in a small number of independent channels. The data were collected by a vertical array designed at the Woods Hole Oceanographic Institution. Phase-shift Keying (PSK) *m*-sequence modulated signals with different carrier frequencies were transmitted at a distance 19.2 km from the array. Even during a strong internal waves activity a low bit error rate was achieved.

PACS numbers: 43.30.Bp [WC]
Date Received: January 22, 2008 **Date Accepted:** April 4, 2008

1. Introduction

Two principle factors that limit the performance of adaptive coherent underwater acoustic communications systems are the ability to adapt the system rapidly enough to keep up with environmental fluctuations or platform motion and a lack of adequate signal energy. The former limitation arises in the context of adaptive equalizers that must estimate the channel impulse response and calculate filter coefficients as the environmental conditions change. The rate of adaptation is limited by the number of degrees of freedom (e.g., the number of taps in the channel impulse response or the equalizer filters) that must be adjusted.

A natural framework for addressing these limiting factors in low frequency acoustic communications is that of mode propagation. In the cases where the receiver has access to an array with enough sensors and aperture to resolve the propagating modes [e.g., a fixed base station that receives signals from autonomous underwater vehicles (AUVs)], wideband modal filtering as an initial processing stage can reduce the number of degrees of freedom (e.g., reduce the effective delay spread of the channel impulse response) that need to be adjusted in later adaptive stages and provides a method of coherently combining all of the received propagating energy in the water column. The former advantage is realized for two reasons. The first is that the number of propagating modes and thus the number of modes containing the significant energy is often less than the number of sensors. Therefore, a problem of working on the outputs of N sensors can be reduced to the problem of operating on the wideband modal time series of M modes where $M < N$. The second reason is that the temporal dispersion of the energy propagating in a single mode is always less than or equal to the total temporal dispersion of the channel. The magnitude of the inequality is dependent on the amount of modal coupling in the propagation channel as well as bandwidth of the transmitted signal. The ability to coherently combine all of the propagating energy in the water column is inherent in the use of modal filtering over all of the propagating modes.

The objectives of this research are to process and analyze the data collected during the Shallow Water acoustics experiment 2006 (SW06) to demonstrate the feasibility of broadband mode decomposition as a preprocessing method to reduce the effective channel delay spread and concentrate received signal energy in a small number of independent channels.

2. Experiment conditions

The processed data were collected by a 48-element array called SHARK, designed at the Woods Hole Oceanographic Institution (Newhall *et al.*, 2007). The array system included a 16 channel vertical array (VLA) and a 32 channel horizontal array. The system was deployed in 78 m of water, which allowed 14 of the 16 vertical array channels to span the water column from about 78 m to 12 m depth. The hydrophone depths (not equidistant) are: 13.5 (top), 17.25, 21, 24.75, 28.5, 32.25, 36, 39.75, 43.5, 47.25, 54.75, 62.25, 69.75, 77.25, 77.75, and 77.75 m.

For broadband mode processing, the first 14 hydrophones were used. The following signals, emitted from the Miami sound machine (MSM) sound source (Newhall *et al.*, 2007) along the bottom slope from the distance 19.2 km, have been processed: Periodically repeated 63 digit *M*-sequence signal with the carrier frequency 101.7 Hz; 127 digit *M*-sequence signal with the frequency 203.4 Hz; 511 digit *M* sequence with frequency 813.8 Hz. The signal was phase modulated with 180° phase shift, four carrier cycles per digit, 40 *M*-sequence periods were processed. The phase shifts take place at zero crossings.

The sound velocity was measured by ten sensors attached to the VLA at the depths: 13, 14.8, 18.5, 22.3, 26, 33.6, 41, 56, 71, and 78.3 m. The sound velocity was measured every 30 s. The mode processing used an averaged value of the sound velocity profile for the time interval covering the signal length.

3. Broadband mode filtering

A number of papers have been dedicated to shallow water acoustic mode problems. Acoustic normal mode filtering and mode fluctuations in shallow water experiments were considered in the paper by Headrick *et al.* (2000). The comparison of different approaches to mode coefficients estimation was conducted in the papers by Buck *et al.* (1998), Wage *et al.* (2003), and Wage *et al.* (2005). In accordance to referenced sources the pseudo-inverse transformation is a very reasonable method for mode coefficient estimation when signal to noise ratio is much larger than one. The first results of mode fluctuations in SW06 were considered by Duda *et al.* (2007), and Collis (2007).

The broadband mode filtering was implemented as follows. The input signal for each hydrophone was sampled and divided into sections. Each section for each hydrophone was zero padded and then Fourier transformed. For each frequency, the Moore–Penrose pseudo inverse of the sampled mode shape matrix was calculated. The pseudo-inverse matrix computation was based on singular value decomposition with large singular values. At each frequency, this inverse was used to perform the modal filtering and the resulting filtered signals were transformed to low frequencies and then inverse Fourier transformed to obtain the modal time series for each mode. The processed sections then were down sampled and combined by overlap-add method. The mode filtering was followed by a correlation with the initial *M*-sequence phase-manipulated signal, and signal demodulation analysis. Another approach based on the low-pass equivalent method (base-band mode filtering) was tested and gave the equivalent result with less computation.

The calculation of the mode shapes used for the mode filtering used data from sound velocity sensors averaged over signal transmission interval. Bottom sound velocity, density, and attenuation were taken from from the paper by Yong-Min Jiang *et al.* (2007). A standard Kraken algorithm was applied for each frequency in the signal bandwidth to calculate frequency dependent acoustic modes and mode wave numbers. Each *m*-sequence signal has four carrier periods in one bit and its bandwidth was $2\Delta f = 0.25 f_0$, where f_0 is the carrier frequency. The frequency dependent mode functions were calculated for the whole signal frequency bandwidth.

4. Correlation of mode filter output with the *M*-sequence signal

The correlation of the signal after the mode filter with the cyclically time-shifted replica of transmitted *M*-sequence phase manipulated signal shows the delay spread of acoustic energy. Correlation of the received signal with the reference signal with a shifted carrier frequency shows that there was no relevant frequency Doppler shift of the received signals. The results for

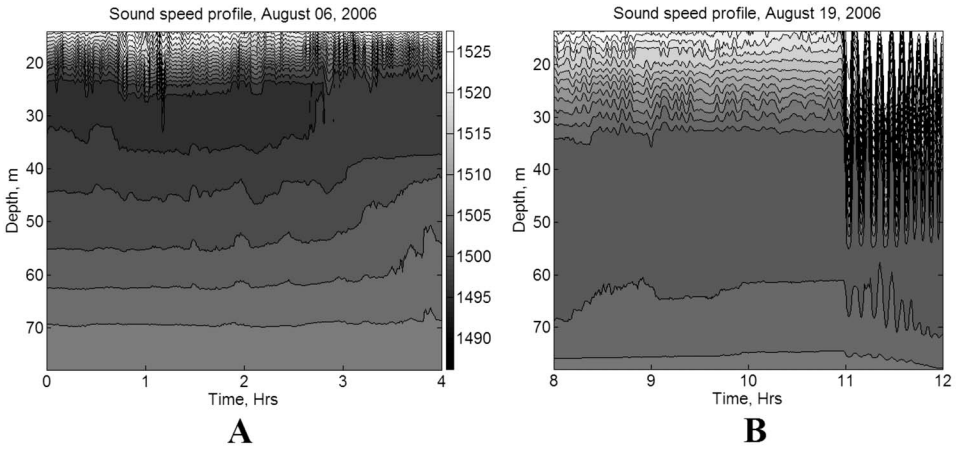


Fig. 1. Sound speed fluctuations.

two different cases are presented below. The first case is "typical." It is a usual internal wave activity, which was observed during most of the SW06 experiment. The second case has "extremely strong" internal wave activity, which leads to strong fluctuations of channel parameters. The last case is not typical, it has been observed only once during the SW06 experiment. These two examples are presented in Fig. 1: ordinary internal wave fluctuations (A) and extremely strong internal wave solitons (B). A tidal wave presents in Fig. 1(a). The correlation functions for these two cases for the carrier 203 Hz are shown in Fig. 2. The constellations of complex envelopes at the output of the mode filters for mode 2 (case A) and mode 1 (case B) are presented in Fig. 3. Signal constellations are not bad and high quality data reception is probable.

5. Data processing results

The good shape of the signal constellations diagram at the output of the mode filter suggests that good results can be expected from data recovery by a short length Viterbi decoder with a simple channel parameters estimator. A joint channel estimation and data recovery algorithm identical to Morozov *et al.* (2006) was applied for quality of data reception demonstration.

The discrete form of the channel at the output of a mode filter is

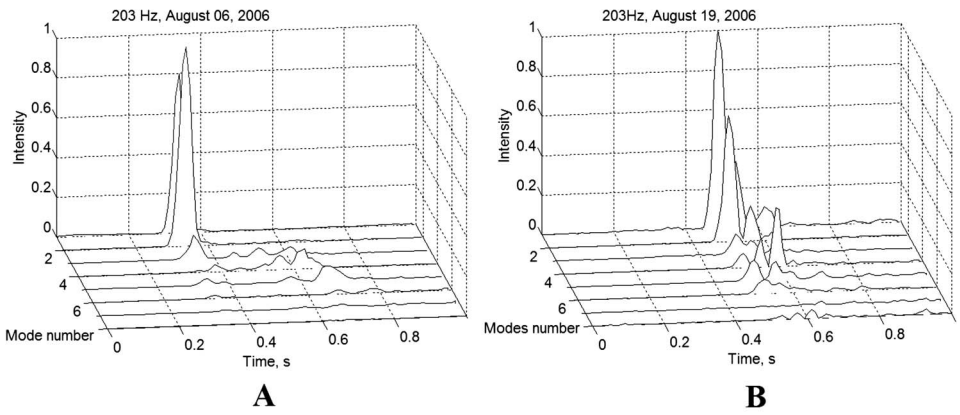


Fig. 2. Mode-time correlations.

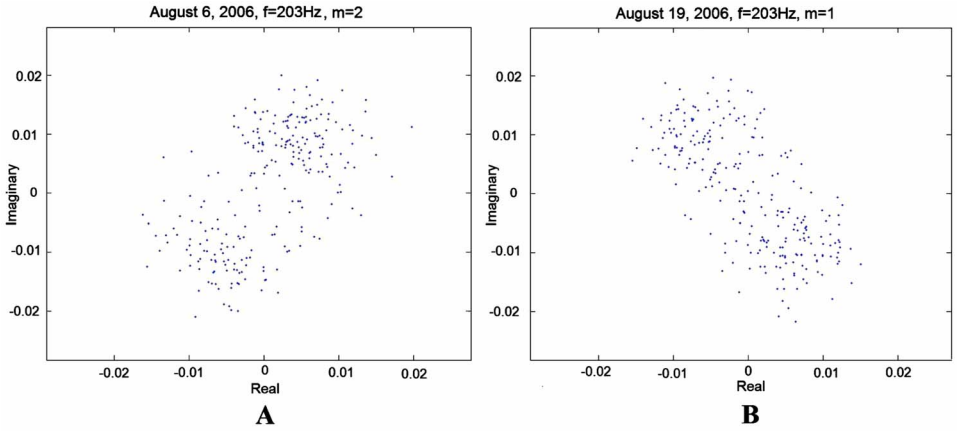


Fig. 3. (Color online) Signal complex envelop constellations at the output of mode filters.

$$y_l = \sum_{j=1}^{N} h_{lj} k_{l-j} + n_l, \tag{1}$$

where n_l is the white Gaussian noise and h_{jl} is the low-frequency equivalent of time-variable channel pulse response; $k_j = \pm 1$ is the modulating parameter.

The simplified equations for the maximum-likelihood metrics w_l are

$$w_l = w_{l-1} - |d_l|^2; \quad \hat{h}_{lm} = \hat{h}_{l-1,m} - a d_l k_{l-m}; \quad d_l = y_l - \sum_{m=0}^{N} \hat{h}_{lm} k_{l-m}, \tag{2}$$

where $\hat{h}_{lm}$ are least mean squares (LMS) estimates of the channel impulse response obtained in Equ. (2); $a = 0.01$ is the step size of the estimation algorithm. The algorithm operates by iterations for each sample with the rate of two samples per pulse. The algorithm calculates trellis metrics in accordance with Eq. 2 and includes an LMS adaptive channel estimator for each sequence of data symbols (for each branch of the trellis).

A receiver estimated mode-time intensity distribution and used the most energetic modes for demodulation. For case A, the data processing gave a bit error rate (BER) of 0.2, when the signal from the first mode filter was received, the BER = 0 for the second mode and 0.2 for the third mode. For case B and the first mode, BER was 0.2, BER = 0 for the second mode, and 0.49 for the third mode. In both cases the second mode filtering gave the best reception quality with no errors.

These results demonstrate that even in a very complicated environment with strong internal waves solitons the data can be transmitted over the distance 19.2 km without errors. Acoustical energy was concentrated in a small number of the first acoustical modes. Note that the sound pressure level of the transmitter was limited to approximately 186 dB.

6. Conclusion

Broadband mode filtering allows one to decompose the received signal into a number of independent signals with a reduced channel delay spread. The data processing shows that even in a very complicated environment with strong internal wave solitons the acoustical energy is concentrated in a small number of the first acoustical modes. A receiver can estimate mode-time intensity distribution and use a signal from a more intensive mode (or a few of them) for demodulation. A very high quality data transmission can be achieved for a range of approximately 20 km.

Acknowledgments

The authors express sincere thanks to Dr. James Lynch and Arthur Newhall for SW06 experimental data. The authors deeply thank Dr. Timothy Duda, Dr. Jon Collis, Dr. Ying Tsong Lin for the help in data processing, Keith Von Der Heydt for the SHARK acoustic array design, and Dr. Harry DeFerrari for the MSM underwater acoustic sound source. The research was supported by ONR.

References and links

Buck, J. R., Preisig, J. C., Wage, K. E., (**1998**). "A unified framework for mode filtering and the maximum *a posteriori* mode filter," J. Acoust. Soc. Am. **103**, 1813–1824.

Collis, J. M., Duda, T. F., Lynch, J. F., and DeFerrari, H. A., (**2008**). "Observed limiting cases of horizontal field cohrerence and array performance in a time-varying internal wave- field," J. Acoust. Soc. Am. **124**, EL97–EL103.

Duda, T. F., and Collis, J. M., (**2007**) "Acoustic field coherence in four-dimensionally variable shallow water environments: Estimation using co-located horizontal and vertical line arrays," Proceedings of the conference "*Underwater Acoustic Measurements, Technologies and Results*," Crete, Greece.

Headrick, R. H., Lynch, J. F., Kemp, J. N., Newhall, A. E., von der Heydt, K., Apel, J., Badey, M., Chiu, C.-S., Finette, S., Orr, M., Pasewark, B., Turdot, A., Wolf, S., and Tielbuerger, D., (**2000**). "Acoustic normal mode fluctuation statistics in 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220.

Morozov, A. K., and Preisig, J. C., (**2006**). "Underwater acoustic communications with multi-carrier modulation." *Proceedings of Oceans 2006*.

Newhall, A. E., Duda, T. F., Von der Heydt, K., Irish, J. D., Kemp, J. N., Lerner, S. A., Liberatore, S. P., Lin, Y. T., Lynch, J. F., Maffei, A. R., Morozov, A. K., Shmelev, A. A., Sellers, C. J., and Witzell, W., (**2007**). "Acoustic and oceanographic observations and configuration information for the WHOI moorings from the SW06 experiment," Applied Ocean Physics and Engineering (AOP&E), WHOI Technical Reports.

Wage, K. E., Baggeroer, A. B., and Preisig, J. C., (**2003**). "Modal analysis of broadband acoustic receptions at 3515-km range in North Pacific using short-time Fourier techniques," J. Acoust. Soc. Am. **113**, 801–817.

Wage, K. E., Dzieciuch, M. A., Worcester, P. F., Howe, B. M., and Mercer, J. A., (**2005**). "Mode coherence at megameter ranges in North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1565–1581.

Yong-Min, Jiang, Chapman, N. Ross, and Badiey, Mohsen, (**2007**). "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data," J. Acoust. Soc. Am. **121**, 4, 1879–1894.

ACOUSTICAL NEWS—USA

Eaine 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.*

Call for Business Meeting of the Society

Notice: A business meeting of the Acoustical Society of America will be held starting at 3:30 p.m. on November 12th at the Doral Hotel and Resort, 4400 NW 87th Avenue, Miami, FL 33178-2192. All ASA Fellows and Members are urged to attend for the purpose of voting on the approval of a Plan of Merger between the Acoustical Society Foundation and the Acoustical Society of America.

Motion to be presented at the business meeting. The motion concerns approving a Plan of Merger by and between the Acoustical Society of America (ASA), Incorporated and Acoustical Society Foundation (ASF) Inc., both being nonprofit organizations. This motion was approved on December 1, 2007 in New Orleans, Louisiana, by a unanimous vote of the Executive Council of the ASA acting on behalf of the Society and as the sole corporate member of the ASF. The entire "Plan of Merger" is printed below.

Motion. To approve the "Plan of Merger by and between the Acoustical Society of America, Incorporated and Acoustical Society Foundation Inc."

Rationale for the Plan of Merger: The merger of the Acoustical Society Foundation into the Acoustical Society of America will eliminate duplication of efforts related to fundraising activities, maintaining an endowment, and dispensing funds. "The merger of the ASF into the ASA will eliminate duplication of efforts and consolidate operations related to fundraising activities, maintaining an endowment, and dispensing funds." As agreed upon by both organizations in the "Memo of Understanding" between the ASA and ASF, the proposed merger will combine assets and restricted funds, and further ensure all funds transferred from the ASF to the ASA will be used only for the intended purposes for which they were donated. The ASA will change its rules to create an Acoustical Society Foundation Board, which will assume responsibilities for fundraising for the ASA.

Charles E. Schmid
Executive Director

PLAN OF MERGER BY AND BETWEEN ACOUSTICAL SOCIETY OF AMERICA, INCORPORATED AND ACOUSTICAL SOCIETY FOUNDATION, INC.

Under Section 902 of the Not-for-Profit Corporation Law

This Plan of Merger (the "Plan") is entered into and effective on _____ 2008, between Acoustical Society of America, Incorporated, a New York not-for-profit corporation ("ASA"), and Acoustical Society Foundation, Inc., a New York not-for-profit corporation (the "Foundation").

WITNESSETH

WHEREAS, ASA and the Foundation (collectively the "Constituent Corporations") are not-for-profit corporations duly organized and existing under the laws of the State of New York;

WHEREAS, as evidenced by the Memorandum of Understanding (the "MOU") dated December 1, 2007 between ASA and the Foundation, a copy of which is attached as Exhibit A, the Executive Council of ASA and the Board of Trustees of the Foundation (the "Constituent Boards") have determined that it is advisable and in the best interest of each of the Constituent Corporations and their respective members that, subject to the terms and conditions set forth in this Plan, the Constituent Corporations combine into a single not-for-profit corporation through the statutory merger of the Foundation with and into ASA (the "Merger"); and

WHEREAS, the Constituent Boards have, subject to the terms and conditions set forth in this Plan, approved in principal the Merger;

NOW, THEREFORE, in consideration of the representations, warranties, and mutual agreements herein contained, the parties hereby agree that all of the assets and liabilities of the Foundation will be acquired and held by ASA, and subsequent to the Effective Date (as defined below), the business of the Foundation will be conducted through ASA, and that ASA will continue its corporate existence. The terms and conditions of the Merger and the manner of carrying it into effect are as follows:

ARTICLE I.

NAMES OF CONSTITUENT CORPORATIONS AND OF SURVIVING CORPORATION

The names of the Constituent Corporations are Acoustical Society of America, Incorporated and Acoustical Society Foundation, Inc. The surviving corporation will be Acoustical Society of America, Incorporated.

ARTICLE II.

MEMBERSHIP

The membership of each Constituent Corporation is as follows.

1. Membership of ASA. Pursuant to Article III of ASA's Bylaws, the membership of ASA is composed of individuals, corporations and organizations admitted to membership. The membership classifications and qualifications are as follows:

(a) Honorary Fellows: Any individual who has attained eminence in acoustics or who has rendered outstanding service to acoustics.

(b) Fellows: Any Member or Associate who has rendered conspicuous service or made notable contributions to the advancement or diffusion of the knowledge of acoustics or the fostering of its practical applications.

(c) Members: Any individual active in acoustics, who has an academic degree in acoustics or in a closely related field or who has had the equivalent of an academic degree in scientific or professional experience in acoustics.

(d) Associates: Any individual interested in acoustics.

(e) Student Members: Any individual enrolled as a student in an accredited educational institution for one-half time or more.

(f) Sustaining Members: Any individual, corporation or organization contributing annual dues as fixed by the Executive Council of ASA.

Members in all classes of membership may attend general meetings of ASA and submit papers for publication in the ASA Journal or for presentation at meetings. Fellows and Members in good standing are entitled to vote and to hold office in ASA.

The number of members in each class as of 30 September 2007 was as follows:

(a) Honorary Fellows: 4
(b) Fellows: 897
(c) Members: 2182
(d) Associates: 3399
(e) Student Members: 1117
(f) Sustaining Members: 27

2. Membership of the Foundation. Pursuant to Section 2.1 of the Foundation's By-Laws, the sole member of the Foundation is ASA.

ARTICLE III.

TERMS AND CONDITIONS OF PROPOSED MERGER

1. Approval.

(a) This Plan has been submitted to the Constituent Boards for discussion and approval.

(b) This Plan has been submitted by ASA to its members for approval.

 0001-4966/2008/124(3)/1379/3/$23.00

(c) This Plan has been submitted by the Foundation to its member for approval.

2. Surviving Corporation. The Foundation will be merged with and into ASA. ASA will survive the Merger under the name Acoustical Society of America, Incorporated and will continue to be organized and existing under the laws of the State of New York, but the separate existence of the Foundation will cease upon the Effective Date.

3. Effective Date of Merger. The proposed Merger will become effective upon the filing of a Certificate of Merger by the Department of State, unless, prior to delivery of such certificate to the Department of State for filing, it is agreed between the Constituent Boards that a later effective date is advisable, in which event the Certificate of Merger will contain a provision specifying such later effective date as is agreed upon between such Boards, but which date under section 905(a) of the New York Not-for-Profit Corporation Law may not exceed thirty days subsequent to the filing of the Certificate of Merger by the Department of State (the "Effective Date").

4. Membership. Since the sole member of the Foundation is ASA, that sole member will not become a member of ASA. ASA will continue to have members as described in Article II, above.

5. Acoustical Society Foundation Fund and Acoustical Society Foundation Board. In accordance with and as described in the MOU, after the Effective Date, ASA will identify certain assets on the ASA books as constituting a fund, which will be designated the Acoustical Society Foundation Fund (the "Fund"). To assist the Executive Council of ASA with decisions regarding the use of assets held in the Fund, maintaining an endowment, fund raising activities, and the granting of funds, the Executive Council of ASA will form a committee of the corporation designated as the Acoustical Society Foundation Board (the "ASF Board"). The ASF Board will be formed and will operate as described in the MOU.

6. Further Assurance of Title. If at any time ASA considers or is advised that any acknowledgements or assurances in law or other similar actions are necessary or desirable in order to acknowledge or confirm in and to ASA any right, title or interest of the Foundation held immediately prior to the Effective Date, the Foundation and its proper officers and directors will execute and deliver all such acknowledgements or assurances in law and do all things necessary or proper to acknowledge or confirm such right, title or interest in ASA as is necessary to carry out the purposes of this Agreement, and ASA and the proper officers and directors thereof are fully authorized to take any and all such action in the name of the Foundation or otherwise.

Rights and Liabilities of ASA. At and after the Effective Date the following will occur:

(a) ASA will succeed to and possess, without further act or deed, all of the rights, privileges, powers, and franchises, both public and private, and all of the property, real, personal, and mixed, of each of the Constituent Corporations.

(b) All debts due to the Foundation on whatever account will be vested in ASA, and all claims, demands, property, rights, privileges, powers, and franchises and every other interest of either of the Constituent Corporations will be as effectively the property of ASA as they were of the respective party hereto.

(c) The title to any real estate vested by deed or otherwise in the Foundation will not revert or be in any way impaired by reasons of the Merger, but will be vested in ASA.

(d) All rights of creditors and all liens upon any property of either of the Constituent Corporations will be preserved unimpaired, limited in lien to the property affected by such lien at the Effective Date; all debts, liabilities, and duties of the Constituent Corporations will thenceforth attach to ASA and may be enforced against it to the same extent as if such debts, liabilities, and duties had been incurred or contracted by it.

(e) All corporate acts, plans, policies, agreements, approvals, and authorizations of ASA and the Foundation, and their members, governing boards and committees thereof, officers and agents, which were valid and effective immediately prior to the Effective Date, will be taken for all purposes as the acts, plans, policies, agreements, approvals, and authorizations of ASA and will be as effective and binding thereon as the same were with respect to either of ASA or the Foundation.

ARTICLE IV.

CERTIFICATE OF INCORPORATION OF SURVIVING CORPORATION

The Certificate of Incorporation of the surviving corporation will be the Certificate of Incorporation of ASA, without amendment or change.

ARTICLE V.

MISCELLANEOUS PROVISIONS

1. Abandonment of Plan. Notwithstanding authorization of this Plan by the Constituent Boards and the members of the Constituent Corporations, if at any time prior to the filing of a Certificate of Merger by the Department of State it becomes the opinion of the governing board of either of the Constituent Corporations that events or circumstances have occurred which render it inadvisable to consummate the Merger, this Plan of Merger may be abandoned. The filing of a Certificate of Merger by the Department of State will conclusively establish that this Plan has not been abandoned.

2. Expenses of Merger. ASA will pay all the expenses of carrying this Plan into effect and of accomplishing the Merger, provided that if at any time this Plan should become abandoned, the Foundation will reimburse ASA for fifty percent of all expenses incurred and paid under this paragraph.

3. Counterparts. For the convenience of the parties and to facilitate approval of this Plan, any number of counterparts of this Plan may be executed and each such executed counterpart will be deemed to be an original instrument.

IN WITNESS WHEREOF, this Plan of Merger has been subscribed on behalf of ASA on _____, 2008, and on behalf of the Foundation on _____, 2008, by the undersigned duly authorized officers of those corporations, the Plan having been duly adopted by the Executive Council of ASA on December 1, 2007, by the members of ASA on _____, 2008, by the Board of Trustees of the Foundation on_, 2008, and by the sole member of the Foundation on December 1, 2007.

ACOUSTICAL SOCIETY OF AMERICA, INCORPORATED	ACOUSTICAL SOCIETY FOUNDATION, INC.
By: _____	By: _____
Name:	Name:
Title:	Title:

USA Meetings Calendar

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

2008

10–14 Nov	156th Meeting of the Acoustical Society of America, Miami, FL [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576- 2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
18–22 May	157th Meeting of the Acoustical Society of America, Portland, OR [Acoustical Soiety of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

Cumulative Indexes to the Journal of the Acoustical Society of America

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division,

American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

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 156th ASA meeting will take place in Miami, Florida, at the Doral Golf Resort and Spa, 10–14 November 2008. More information, including online registration, hotel information and a Call for Papers, can be found on the ASA website at http://asa.aip.org/meetings.html.

• ***Monday, November 10, 2008***

7:00–10:00 p.m.—ASACOS Steering Committee

• ***Tuesday, November 11, 2008***

—7:30–9:00 a.m.—ASA Committee on Standards (ASACOS) Meeting of the Committee that directs the Standards Program of the Acoustical Society

—9:15–10:30 a.m.—Standards Plenary Group (includes U.S. TAGs to ISO/TC 43, TC 43/SC 1, ISO/TC 108 and its subcommittees, and TC 29)

—11:00 a.m.–12:45 p.m.—Accredited Standards Committee S12, Noise

—2:15–3:30 p.m.—Accredited Standards Committee S1, Acoustics

—3:45–5:00 p.m.—Accredited Standards Committee, S2, Mechanical Vibration and Shock

• ***Wednesday, November 12, 2008***

—9:00–10:30 a.m.—Accredited Standards Committee, S3, Bioacoustics

—10:45 a.m.–12:00 noon—Accredited Standards Committee S3/SC1, Animal Bioacoustics

Standards Meetings Calendar International

• ***8–12 September 2008***

ISO/TC108/SC 4 Human exposure to mechanical vibration and shock will meet in Oslo, Norway.

• ***3–7 November 2008***

ISO/TC 108 Mechanical vibration, shock and condition monitoring, will meet in St. Louis, Missouri, USA, along with subcommittee **ISO/TC 108/SC 3** Use of calibration of vibration and shock measuring instruments. **ISO/TC 108/SC 6** will also meet in St. Louis, Missouri, **4–6 November 2008**.

• ***30 March–3 April, 2009***

ISO/TC 108/SC 2 Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures will meet in London, UK.

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 withdrawal approval of existing standards. The dates listed in parenthesis are for information only.

ASA (ASC S2) (Acoustical Society of America)

New Standards

BSR/ASA S2.62-200x, Shock Test Requirements for Equipment in a Rugged Shock Environment (new standard)

To be used for testing equipment that will be subjected to shock. Defines test requirements and severity thresholds for a large range of shock environments, including but not limited to shipping, transport, and rugged operational environments. This standard will allow vendors to better market, and users to more easily identify, equipment that will operate or simply survive in rugged shock environments. Standard includes references to various ASTM, IEC, NATO, and US military standards. (August 4, 2008)

ASA (ASC S12) (Acoustical Society of America)

Revisions

BSR/ASA S12.6-200x, Methods for Measuring the Real-Ear Attenuation of Hearing Protectors [revision and redesignation of ANSI S12.6-1997 (R2002)]

Specifies lab-based procedures for measuring, analyzing and reporting passive noise-reducing capabilities of hearing protection devices. Procedures consist of psychophysical tests conducted on human subjects to determine real-ear attenuation measured at hearing threshold. (August 18, 2008)

BSR/ASA S12.9-Part 6-200x, Quantities and Procedures for Description and Measurement of Environmental Sound—Part 6: Methods for Estimating of Awakenings Associated with Outdoor Noise Events Heard in Homes (revision and redesignation of ANSI S12.9—Part 6-2000 (R2005))

Provides a method to predict sleep disturbance in terms of percent awakenings or numbers of people awakened associated with noise levels in terms of indoor A-weighted sound exposure level (ASEL). Developed from field studies of behavioral awakening primarily in homes near areas of routine jet aircraft takeoff and landing operations, railroads, roads, and highways. The database used in derivation of the method consists of about 10 000 subject-nights of observations in a variety of communities. (June 30, 2008)

Reaffirmations

BSR/ASA S12.11-2003 Part 1/ISO 10302:1996 (MOD) (R200x), Acoustics—Measurement of noise and vibration of small air-moving devices, Part 1: Airborne noise emission (a Modified NAIS) [reaffirmation and redesignation of ANSI S12.11/1 ISO 10302-1996 (MOD)-2003]

Specifies in detail a laboratory method for determining and reporting airborne noise emissions of small air-moving devices used primarily for cooling electronic equipment like computer and business equipment. Descriptor of overall noise emission of air-moving device under test is A-weighted sound power level. 1/3-octave-band sound power level is the detailed descriptor of noise emission. Octave band sound power levels may be provided in addition to 1/3-octave-band sound power levels. (June 30, 2008)

BSR/ASA S12.11-2003/Part 2 (R200x), Acoustics—Measurement of noise and vibration of small air-moving devices, Part 2: Structure-borne vibration (reaffirmation and redesignation of ANSI S12.11/2-2003)

Contains the recommended methods for testing, determining, and reporting the vibration levels induced by small air-moving devices (AMDs) that are found in cooling equipment used for information technology and telecommunications. The use of this standard is encouraged to promote uniformity in the measurement and reporting of the vibration levels induced by AMDs for use in information technology and telecommunications equipment. (June 30, 2008)

ABMA (ASC B3) (American Bearing Manufacturers Association)

Reaffirmations

BSR/ABMA 13-1987 (R200x), Rolling Bearing Vibration and Noise (Methods of Measuring) [reaffirmation of ANSI/ABMA 13-1987 (R1999)]

Serves to define and specify, for purposes of bearing quality assurance, the physical quantities measured and the test conditions utilized in measurement of vibration and noise generated by roller bearings. (August 12, 2008)

NSF (NSF International)

Revisions

BSR/NSF49-200x (i19), Class II (laminar flow) biosafety cabinetry (revision of ANSI/NSF 49-2007)

Issue 19—Revise Noise Level Test acceptance criteria to be consistent with background noise levels. (June 15, 2008)

UL (Underwriters Laboratories, Inc.)

Revisions

BSR/UL 1971-200x, Signaling Devices for the Hearing Impaired (revision of ANSI/UL 1971–2004)
Revises the ratings information in installation literature. (June 29, 2008)

BSR/UL 60745-2-12-200x, Standard for Safety for Hand-Held Motor-Operated Electric Tools—Safety—Part 2-12: Particular Requirements for Concrete Vibrators (revision of ANSI/UL 60745-2-12-2005)
Provides: (1) Proposed revisions to align with IEC Amendment No. 1 for IEC 60745-2-12; and (2) Addition of a national difference to Clause 24.4 to require hard service type S cord, that is oil and weather resistant. (July 28, 2008)

Call for Members (ANS Consensus Bodies)

Directly and materially affected parties who are interested in participating as a member of an ANS consensus body for the standards listed below are requested to contact the sponsoring standards developer directly and in a timely manner.

ASA (ASC S2) (Acoustical Society of America)

BSR/ASA S2.74-200x, Guide for the Measurement and Health and Comfort Evaluation of Human Exposure to Whole-Body Vibration (new standard)

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.

ASA (ASC S12) (Acoustical Society of America)

BSR/ASA S12.60-200x/Part 3-200x, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools—Part 3: Information Technology Equipment in Classrooms (new standard)

Develops a new part to ANSI S12.60 to cover IT equipment with (1) recommended sound power criteria for IT equipment in classrooms; (2) recommendations for purchase specification format; and (3) recommendations for installation conditions of IT equipment. Project Need: To inform school administrators, purchasing agents, school architects and designers, teachers and parents about the problem of noise caused by portable classroom equipment (including IT equipment), and the solutions to the problem. Stakeholders: School administrators, purchasing agents, school architects and designers, teachers and parents.

BSR/ASA S2.74-200x, Guide for the Measurement and Health and Comfort Evaluation of Human Exposure to Whole-Body Vibration (new standard)

Specifies the recommended method for the measurement, data analysis and evaluation, health risk and comfort assessments, and reporting of whole-body vibration (WBV) exposure. The methods specified in this standard are to be used to characterize and evaluate vibration impinging on the human body through contact with a supporting surface. Guidance is provided for assessing health risk and comfort related to WBV exposure. Project Need: To provide information about the risks associated with exposure to whole-body vibration or to provide guidance on assessing the potential

for such exposure to result in pathology. Stakeholders: Medical community, engineers, employers whose employees are subjected to mechanical vibration in the workplace.

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. Technical Reports have been registered in accordance with ANSI's *Procedures for the Registration of Technical Reports with ANSI*.

ASA (ASC S1) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S1.14-1998 (R2008), Recommendations for Specifying and Testing the Susceptibility of Acoustical Instruments to Radiated Radio-Frequency Electromagnetic Fields [reaffirmation and redesignation of ANSI S1.14-1998 (R2003)]: 4/25/2008

ASA (ASC S2) (Acoustical Society of America)

Revisions

ANSI/ASA S2.9-2008, Parameters for Specifying Damping Properties of Materials and System Damping [revision and redesignation of ANSI S2.9-1976 (R2006)]: 6/13/2008

IEEE (Institute of Electrical and Electronics Engineers)

Reaffirmations

ANSI/IEEE 664-1994 (R2007), Guide for Laboratory Measurement of the Power Dissipation Characteristics of Aeolian Vibration Dampers for Single Conductors [reaffirmation of ANSI/IEEE 664-1994 (R2000)]: 4/24/2008

Newly Published ISO and IEC Standards

Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization

ISO Standards

ACOUSTICS (TC 43)

ISO 3382-2:2008, Acoustics—Measurement of room acoustic parameters—Part 2: Reverberation time in ordinary rooms

MACHINE TOOLS (TC 39)

ISO 8525:2008, Airborne noise emitted by machine tools—Operating conditions for metal-cutting machines

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 18431-2/Cor1:2008, Mechanical vibration and shock—Signal processing—Part 2: Time domain windows for Fourier Transform analysis—Corrigendum

ISO 18437-4:2008, Mechanical vibration and shock—Characterization of the dynamic mechanical properties of visco-elastic materials—Part 4: Dynamic stiffness method

IEC Standards

ENVIRONMENTAL CONDITIONS, CLASSIFICATION AND METHODS OF TEST (TC 104)

IEC 60068-2-64 Ed. 2.0 b:2008, Environmental testing—Part 2-64: Tests—Test Fh: Vibration, broadband random and guidance

ISO Draft International Standards

This section lists proposed standards that the International Organization for Standardization (ISO) is considering for approval. The proposals have received substantial support within the technical committees or subcommittees that developed them and are now being circulated to ISO members for comment and vote. Standards Action readers interested in reviewing and commenting on these documents should order copies from ANSI.

ACOUSTICS (TC 43)

ISO 7779/DAmd2.2, Revision of measurement surfaces, procedures for equipment installation/operation and detection of prominent discrete tones—6/6/2008

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 16063-31, Methods for the calibration of vibration and shock transducers—Part 31: Testing of transverse vibration sensitivity—8/3/2008

BOOK REVIEWS

P. L. Marston
Physics Department, Washington State University, Pullman, Washington 99164

These reviews of books and other forms of information express the opinions of the individual reviewers and are not necessarily endorsed by the Editorial Board of this Journal.

Transducers and Arrays for Underwater Sound

Charles H. Sherman and John L. Butler

Springer: Monograph Series in Underwater Acoustics, 2007. 610 pp. Price $149.00 (hardcover), ISBN: 9780387329406

The Office of Naval Research began a program in 2001 to fund the preparation and publication of selected monographs in the field of underwater acoustics. The goal was to create a set of in-depth reviews on our current state of understanding the physics of sound in focused research areas including high-frequency seafloor acoustics, matched field processing, acoustical oceanography of bubbles, ocean ambient noise, shallow water acoustics, and underwater transducers and arrays.

Charles H. Sherman and John L. Butler, prominent researchers in the field who have worked in government, corporate and small engineering research and development firms, were selected to prepare the monograph ***Transducers and Arrays for Underwater Sound***. This may be one of the more challenging assignments in the underwater series due to the extensive breath and depth involved in this highly multidisciplinary field which includes physical acoustics, transduction science, electrical science, material science, chemistry and crystallography, continuum mechanics, mechanical engineering, acoustical engineering, electrical engineering, signal optimization and noise reduction, analytical and finite element modeling, metrology, test and calibration, as well as related design, fabrication, production, and reliability engineering challenges.

This new book is a welcome contribution to a field in which there are very few monographs or texts, due in part to limitations placed on the distribution of underwater transducer information for more than a half century. Material that was viewed as potentially classified or sensitive by the cognizant security officer, or competitively "useful" by corporations that guard the release of almost everything, were often restricted. The few previous monograph/texts in the area of underwater electroacoustic transducers that did find their way through the government and industrial collaboration are worth mentioning and include: *Electroacoustics*, by Frederick Hunt (J. Wiley & Sons: 1954, pp. 260), *Introduction to Theory and Design of Sonar Transducers*, by O. Bryan Wilson (Peninsular Press, Los Altos: 1985, pp. 191), *Sonar Transducer Fundamentals*, by Ralph Woollett (NUWC, Newport: undated but c. 1986, pp. 102, course notes published posthumously after a tragic accident that claimed the life of the foremost U.S. authority of his time), *Electromechanical Piezoceramic Transducers*, by Boris Aronov (Energoatomizdat, Leningrad [St. Petersburg, Russia]: 1990, pp. 271 and translated in 2007–2008), and *Underwater Electroacoustic Transducers*, by Dennis Stansfield (Bath Univ. Press, Bath, UK, 1991, pp. 413).

In comparison, the 610 page *Transducers and Arrays for Underwater Sound* is more comprehensive, containing a much broader scope, and providing coverage in areas including: Introductory treatment of transducers and arrays, in-depth coverage of several types of the most common transduction approaches (piezoelectric, electrostrictive, magnetostrictive, electrostatic, variable reluctance, and moving coil magnetic transduction), examples of individual projectors and hydrophones and their electrical circuit analogs, an introduction to arrays of underwater projectors and hydrophones, array properties and characterization, highlights of recent developments in the field, summaries of complications related to the design and implementation of arrays including mutual interaction between elements, noise and in its many origins. The book also includes chapters on the introduction to analytical and numerical modeling techniques, nonlinear effects in transducers (some desirable and some unwanted effects), the mechano-acoustic radiation problem, summaries of transducer measurement and calibration, and perspectives for the direction and growth of the field. The appendices provide a broad offering of useful information conveniently assembled for easy reference, such as an expansive list of the properties of traditional as well as new transduction materials such as single-crystal relaxor ferroelectrics. Separately, the authors have also compiled a collection of home problems (exercises) and solutions that may be accessed electronically through the publisher's website (springer.com) so that the monograph could potentially be used as a text for upper level undergraduate or graduate study.

There are many references to both classic and contemporary works, however several are listed as just private communications or internal company reports that are not available to the public, and thus are included as a matter of credit and do not provide additional material that would be easily accessible for the interested reader. There were several missed opportunities to include particularly relevant references in both the U.S. and foreign literature.

Whether the book is viewed as a technical monograph, a review of published works, a multidisciplinary guide for the practitioner, or as a graduate text is left to the reader; it is clearly a welcome contribution to the field of transducers and arrays, as well as the broader field of underwater acoustics. It may also be useful for those involved in related fields such electromechanical actuator design, nondestructive evaluation, medical and biological ultrasonics or in assisting with the training of a new generation of electroacoustic scientists and engineers.

DAVID A. BROWN
University of Massachusetts Dartmouth
Professor of Electrical Engineering and Adjunct Professor of Physics, Director: Electro-Acoustics Research Laboratory
Dartmouth, Massachusetts 02743
and
President
BTech Acoustics LLC
17 Surrey Road
Barrington, RI 02806

Structure of materials: An introduction to crystallography, diffraction, and symmetry

Marc De Graef and Michael E. McHenry

Cambridge University Press, Cambridge, 2007. 844 pp. Price $95.00 (hardcover), ISBN: 978-0-521-65151-6.

For the past few years, I have been searching for an adequate textbook for my graduate-level class "Introduction to materials science: Structure, bonding, and diffraction," without much satisfaction. Every year, I end up using one of the many undergraduate textbooks in materials science that are available, and supplement with more detailed material gathered from various sources. I was very pleased, therefore, to discover the publication of a new textbook, "*Structure of materials, an introduction to crystallography, diffraction, and symmetry*." This book will indeed put an end to my quest, as it provides an exhaustive coverage of crystallography, symmetry, and diffraction, suitable for a one-semester graduate (or advanced undergraduate) course in materials science, physics, or chemistry.

What I particularly like about the book is its choice of topics that are covered and the way these topics are organized. The first half provides an in-depth treatment of symmetry and crystallography, providing all the tools necessary to describe crystal structures in an unambiguous way. Several concepts of diffraction (including structure factor) are introduced at the end of the first part, in chapters 11–13. The second half of the book discusses the

 0001-4966/2008/124(3)/1/0/$23.00

structures of several classes of materials, covering metals, oxides, and molecular solids. Both complex atomic structures as well as technologically important materials are illustrated. Sections and illustrations of this part can easily be used to illustrate various concepts of the first part. Some of the structures discussed are quite advanced, but these can simply be omitted in an introductory class without losing the general thread of the book. The link between both parts of the book is provided in Chapter 14, where the techniques introduced in the first chapters are used to study materials discussed in the second part.

The various crystal structures are easily visualized through clear illustrations. A particularly nice feature is that these images are also available on the book's website, rendered in color, along with some basic structure information. High resolution animated GIF movies of the 32 point groups, as well as JPEG images of each point group, can also be downloaded from the book's website. Another nice touch is the "historical note" at the end of each chapter, providing the reader some background information by reflecting on important events in the history of crystallography and/or important scientists that have made a contribution to the field. Chapter 14, for instance, reproduces the paper by father and son Bragg on the structure of diamond.

While the book is quite long (over 800 pages), and the authors have addressed a wide range of topics on structure and diffraction, there is at least one topic I would have liked to see expanded: A discussion on defects in materials. Two pages are dedicated to defects in ceramics, which seems rather scanty to me. An additional chapter discussing e.g. point and line defects, as well as twinning, would be a nice addition to this otherwise very comprehensive textbook.

Does the book have any shortcomings? Sure. Every chapter comes with an extensive set of exercises (solutions for the instructor are available from the publisher), dealing with the concepts introduced in that chapter. But there is no "answer to selected problems" section, which I typically find a useful confidence builder for the reader who likes to find out if he "got it right." Some chapters partially make up for this by containing a "worked examples" section, guiding the reader through the solution of a problem, but an appendix with a list of answers to some of the problems would be a supplement worth considering for possible new editions. In addition, the book contains at least one rather irritating hiccup: Ruddlesden-Popper is consistently misspelled as "Ruddelson-Popper," a somewhat shameful glitch in an otherwise well-edited book.

Who will benefit from this book? The primary audience would be students (and instructors) in materials science, physics, and/or chemistry, who will find it a very comprehensive textbook on crystallography, symmetry, and diffraction, despite the lack of answers to problems mentioned above. In addition, anybody that wishes to visualize and/or familiarize himself with the structure of a wide variety of metals and ceramics is bound to find the text and its associated website an excellent resource.

In summary, *Structure of materials* is a well-written textbook on crystallography and diffraction. Its choice of topics and abundant illustrations make it well suited for an advanced undergraduate/introductory graduate class in materials science, physics, or chemistry. It covers all the basics while providing ample opportunity to elaborate on more complex topics in the field.

VEERLE KEPPENS
Department of Materials Science and Engineering,
The University of Tennessee,
312 Dougherty Hall,
Knoxville, TN 37996

 0001-4966/2008/124(3)/1/0/$23.00

REVIEWS OF ACOUSTICAL PATENTS

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

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

The purpose of these acoustical patent reviews is to provide enough information for a Journal reader to decide whether to seek more information from the patent itself. Any opinions expressed here are those of reviewers as individuals and are not legal opinions. Printed copies of United States Patents may be ordered at $3.00 each from the Commissioner of Patents and Trademarks, Washington, DC 20231. Patents are available via the internet at http://www.uspto.gov.

Reviewers for this issue:

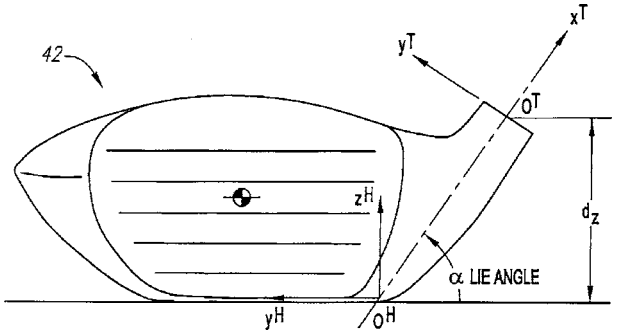
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,320,646

43.20.Tb MULTIPLE MATERIAL GOLF CLUB HEAD

J. Andrew Galloway, assignor to Callaway Golf Company
22 January 2008 (Class 473/224); filed 9 October 2006

Sound modifying component **59** is located on an external surface of golf club face component **60**. "The sound modifying component alters the



frequency, amplitude, and duration of sound made when the golf club head strikes a golf ball." How this may improve one's handicap is not stated.—NAS

7,350,567

43.30.Ma INCREASING MEDIA PERMEABILITY WITH ACOUSTIC VIBRATIONS

Larry G. Stolarczyk and Dimitri A. Kas'yanov, both of Raton, New Mexico
1 April 2008 (Class 166/249); filed 21 October 2005

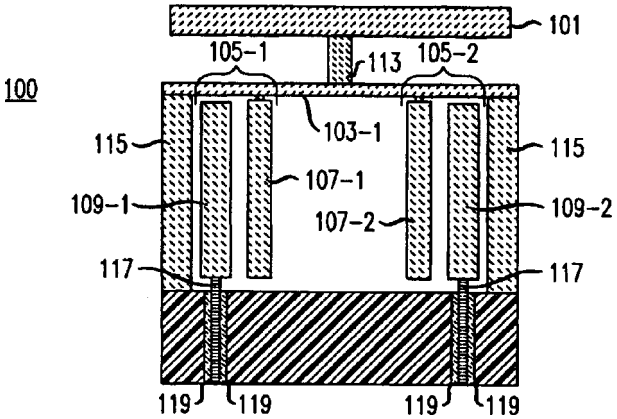
This patent describes the use of vibratory excitation of drill strings to enhance the production of methane from drilled wells. The concept is simple, in that there are no moving parts to the sonic excitation apparatus; it is all accomplished by the use of whistles whose working fluid is water or air. These whistles are located at the perimeter of the drill string and radiate into the surrounding coal bed. The authors provide no measured data to back up their claims that this would work, and do not seem to have accounted for the relatively poor sound propagation in coal beds.—JAH

7,344,262

43.38.Ar MEMS MIRROR WITH TIP OR PISTON MOTION FOR USE IN ADAPTIVE OPTICS

Dennis S. Greywall, assignor to Lucent Technologies Incorporated
18 March 2008 (Class 359/849); filed 29 September 2004

This patent discloses the use of vertically oriented drive electrodes for a microelectronic mechanical system (MEMS) mirror actuator application. The approach is somewhat unconventional due to the fact that all the electrodes are vertical plates, and their connections and support points can be located within a box whose boundaries are defined by the mirror edges on the top of the device, thus saving overall chip area and eliminating dark



space around the mirrors. In the figure, **101** is the mirror, which is hinged along a long and narrow support wall **113**, seen in cross section here. Wall **113** rests on a thin plate that acts as a spring and hinge for tipping plate **101** left and right or pulling it down. Electrodes **109** affect the actuation by pulling the hinge plate electrodes **107-1** and **107-2**. If one is pulled, the mirror tips; if both are pulled, then the mirror retracts and pulls down. This allows the mirrors to be used for adaptive optics, phase shift elements for

 0001-4966/2008/124(3)/1387/13/$23.00

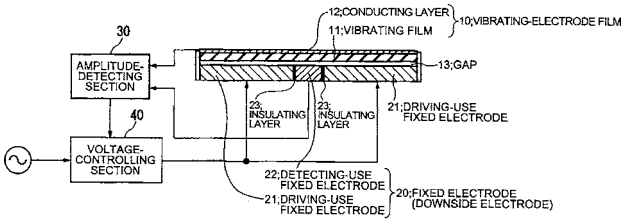
lithography, or projection elements in pixel-addressable reflective optics. Some of the variations described in the patent are probably too complex to contemplate fabricating given the current state of MEMS, but others looked practical.—JAH

7,355,322

43.38.Fx ULTRASONIC TRANSDUCER, ULTRASONIC SPEAKER AND METHOD OF DRIVING AND CONTROLLING ULTRASONIC TRANSDUCER

Shinichi Miyazaki, assignor to Seiko Epson Corporation
8 April 2008 (Class 310/316.01); filed in Japan 18 June 2004

Although piezoelectric transducers are most often used for ultrasonic frequencies this patent argues that an electrostatic transducer offers numerous benefits. An improved single-ended design is described that has an additional sensing electrode to provide motional feedback.—GLA

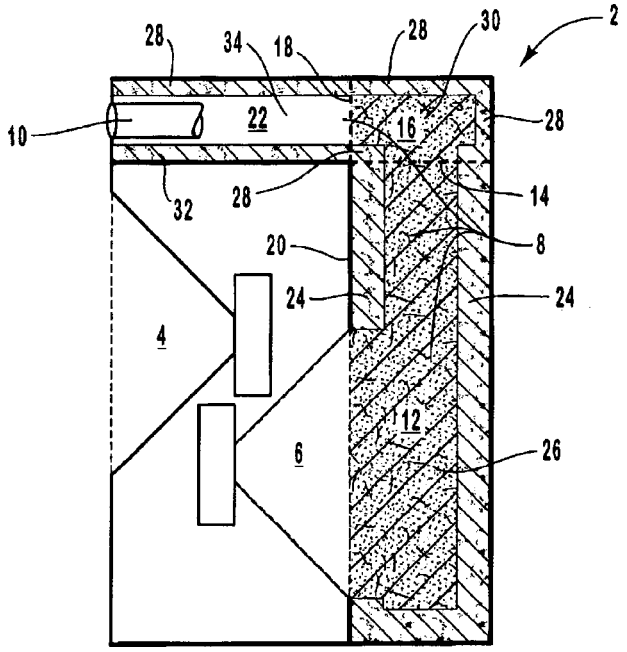


7,325,649

43.38.Ja LOUDSPEAKER WITH PROGRESSIVELY DAMPED ACOUSTICAL CHAMBER

Tierry R. Budge, assignor to Farnsworth & Budge, LLC
5 February 2008 (Class 181/151); filed 23 September 1999

To solve the problem of controlling "driver backwaves and the air pressure and flow at the rear of a driver," as well as the pesky problem of the inertia of a driver when passing though the driver's "neutral" position when it is displaced, a method using a damping material **12**, with a density that decreases with increasing distance from driver **6**, is described. This helps mitigate the assumed problems with radiation from driver **6** "since sound travels more slowly through denser materials." Drivers **4** and **6** also appear to be driven in phase.—NAS

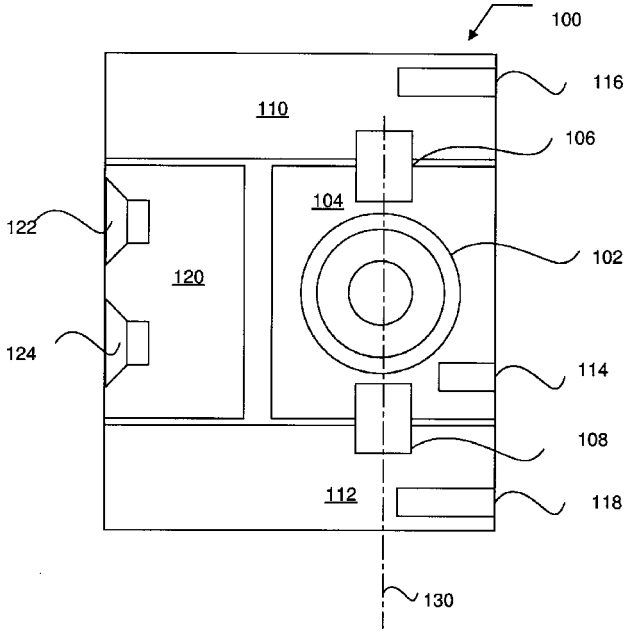


7,350,618

43.38.Ja MULTIMEDIA SPEAKER PRODUCT

Aik Hee Goh *et al.*, assignors to Creative Technology Limited
1 April 2008 (Class 181/156); filed 1 April 2005

Loudspeaker designs that rely on wishful wizardry are still with us. This three-chamber loudspeaker system includes a side-firing woofer and front-firing mid and treble speakers. At this point the design becomes difficult to understand. It apparently includes multiple vents and a variable valve between two of the chambers. "In this way, movement of the main chamber simultaneously causes a constriction or expansion in the aperture, allowing the tailoring of the characteristics of the speaker system to the listening environment."—GLA

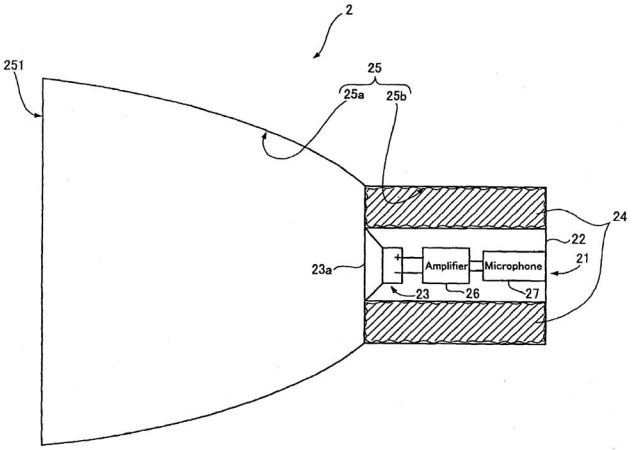


7,352,875

43.38.Ja SPEAKER APPARATUS

Hajime Hatano, Fujisawa-shi, Kanagawa 251-0038, Japan et al.
1 April 2008 (Class 381/160); filed in Japan 12 November 2003

A highly directional loudspeaker system can be nothing more than a small speaker and a parabolic reflector, and such systems are commercially available. This patent describes a somewhat similar arrangement using a reflective hood rather than a parabolic dish. The cavity behind the loudspeaker contains sound absorbing material and may also house electronic equipment.—GLA

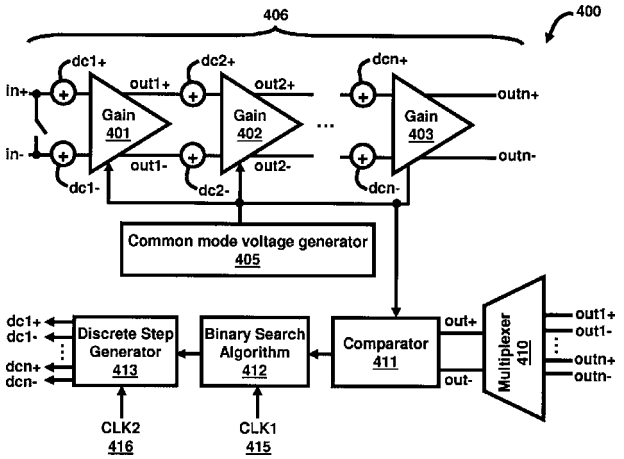


7,348,839

43.38.Lc METHOD AND APPARATUS FOR DC OFFSET CANCELLATION IN AMPLIFIERS

Amr Fahim *et al.*, assignors to Newport Media, Incorporated
25 March 2008 (Class 330/9); filed 23 August 2006

Multiple-stage direct-coupled amplifiers are plagued by dc offset errors, requiring additional error correction circuitry. This patent gives a brief survey of prior art and then describes a digital logic circuit that is said to



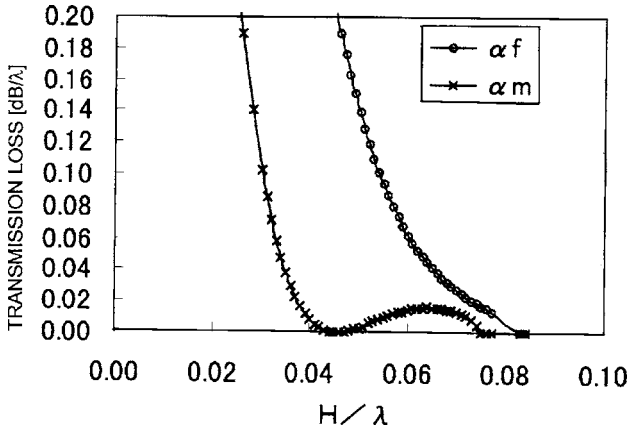
achieve high precision without the drawbacks of other methods. An adaptive search algorithm is combined with fixed and modulated switching in a dc cancellation loop. The calibration circuitry is turned off after the loop has settled.—GLA

7,315,107

43.38.Rh SURFACE ACOUSTIC WAVE DEVICE

Hajime Kando *et al.*, assignors to Murata Manufacturing Company, Limited
1 January 2008 (Class 310/313 R); filed in Japan 12 August 2002

A surface acoustic wave filter is claimed utilizing the second leaky wave that has high velocity and low loss, propagating on the surface of a



lithium niobate or lithium tantalate substrate cut on a plane of special Euler angles.—AJC

7,346,172

43.38.Si AUDITORY ALERT SYSTEMS WITH ENHANCED DETECTABILITY

Durand R. Begault, assignor to The United States of America as represented by the United States National Aeronautics and Space Administration
18 March 2008 (Class 381/56); filed 28 March 2001

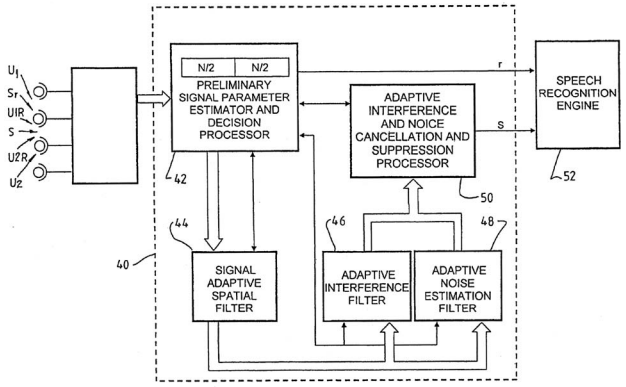
A large amount of work has been devoted to methods for distinguishing a warning signal from other auditory information or background noise. This patent provides a brief summary of prior art, then discloses three new approaches. All three are covered in the abstract, described at length in the patent text, and augmented by illustrations. It turns out however that only one of the three has actually been patented; the other two are not mentioned in the patent claims. The warning signal is presented stereophonically via loudspeakers or headphones such that its perceived location shifts within a relatively short time interval.—GLA

7,346,175

43.38.Si SYSTEM AND APPARATUS FOR SPEECH COMMUNICATION AND SPEECH RECOGNITION

Siew Kok Hui *et al.*, assignors to Bitwave Private Limited
18 March 2008 (Class 381/74); filed in Singapore 12 September 2001

Although the signal processing circuitry described here may be used in conjunction with a speech recognition engine, the patent deals only with a method of processing audio signals received by an array of two or more



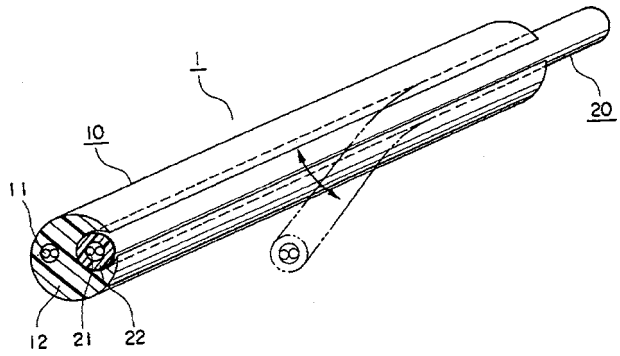
microphones. A first adaptive filter enhances the target signal. Its output is fed to a second adaptive filter that reduces interference signals and a third adaptive filter arranged to reduce noise.—GLA

7,348,490

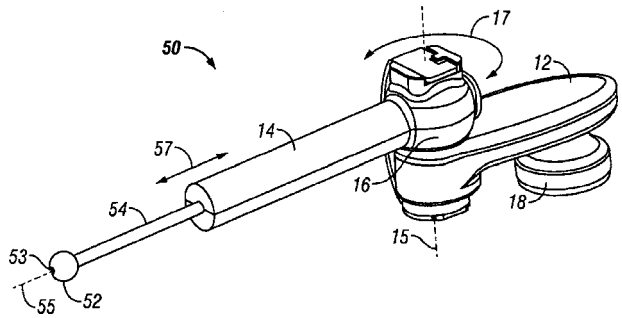
43.38.Si DOUBLE WIRE CORD AND PORTABLE AUDIO DEVICE USING THE SAME

Kiyotaka Sakai, assignor to SDK Company
25 March 2008 (Class 174/113 R); filed in Japan 30 May 2006

Although this patent does not actually deal with acoustical issues, the invention is interesting. Pairs of cables that connect stereo headphones may sometimes be separated, whereas at other times they may be tied together for most of their length. Prior art includes various kinds of clips and sliding



fasteners. In this case the two cables are designed to snap together like a freezer bag.—GLA



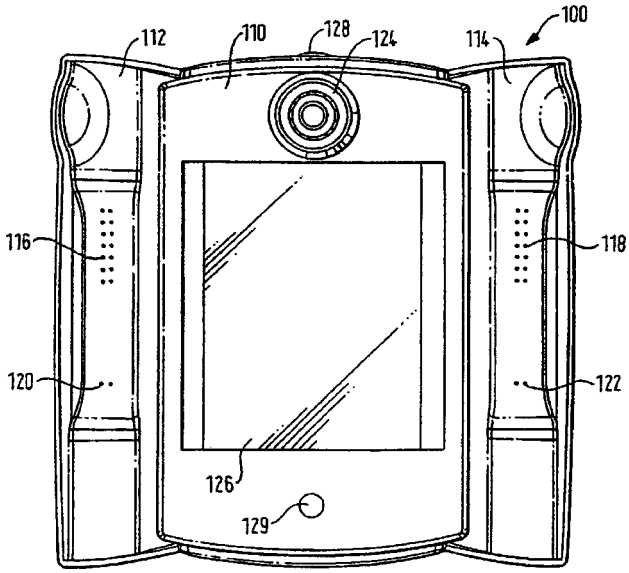
background noise is high. A sensor and controller circuit automatically adjust sensitivity and equalization according to individual positions.—GLA

7,349,004

43.38.Si MOBILE TELEPHONE HANDSET

Paul Priestman *et al.*, assignors to Orange Personal Communications Services Limited
25 March 2008 (Class 348/14.01); filed in United Kingdom 8 May 1998

This multipurpose cellular videophone can also be used for stereo



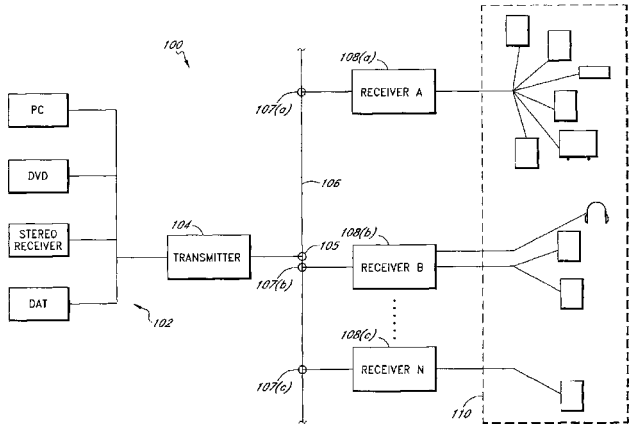
music playback through speakers **116** and **118** that are housed in little fold-out panels **112** and **114**.—GLA

7,346,332

43.38.Vk WIRED, WIRELESS, INFRARED, AND POWERLINE AUDIO ENTERTAINMENT SYSTEMS

William A. McCarty and Yadir Rodriquez, assignors to KSC Industries Incorporated
18 March 2008 (Class 455/402); filed 3 July 2003

Wireless microphones have been available for more than 50 years, but it is now possible to buy wireless headphones, loudspeakers, and audio mixers. This patent (a continuation of a half-dozen earlier patents and applications) deals with a complete home entertainment network. According to



the title, abstract, and patent text the network can be wired, wireless, infrared, or powerline. However, the 60 patent claims are restricted to "an apparatus for receiving an audio signal via a powerline network." Individual (and mostly obvious) control functions are described step by step in seemingly endless detail.—GLA

7,349,547

43.38.Si NOISE MASKING COMMUNICATIONS APPARATUS

Osman K. Isvan, assignor to Plantronics, Incorporated
25 March 2008 (Class 381/104); filed 20 November 2001

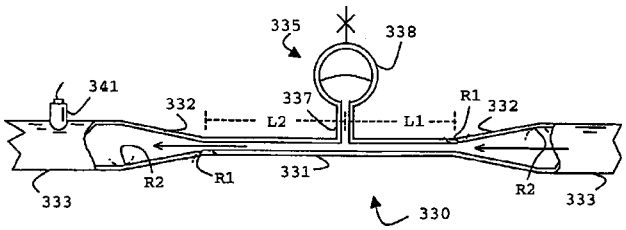
An extensible, swiveling headset microphone boom is described that can be used in two or more positions. For example, a user may normally place the microphone near the cheek, but move it closer to the mouth when

7,345,594

43.40.Sk NOISE ATTENUATION APPARATUS FOR BOREHOLE TELEMETRY

Songming Huang *et al.*, assignors to Schlumberger Technology Corporation
18 March 2008 (Class 340/854.3); filed in United Kingdom 6 September 2002

This patent discloses some of the methods used in the petroleum industry to transmit data from down-hole instruments to the surface, and a specific fluid-mechanical filter designed to reduce the background noise for detection of the telemetry signals. In the drilling process, a continuous stream of data from compasses, densitometers, and pressure sensors is sent to the surface via pressure pulses in the drilling mud that fills the borehole. The present invention describes a couple of related mechanical filters that can be used to separate out and attenuate the upward-traveling data from the



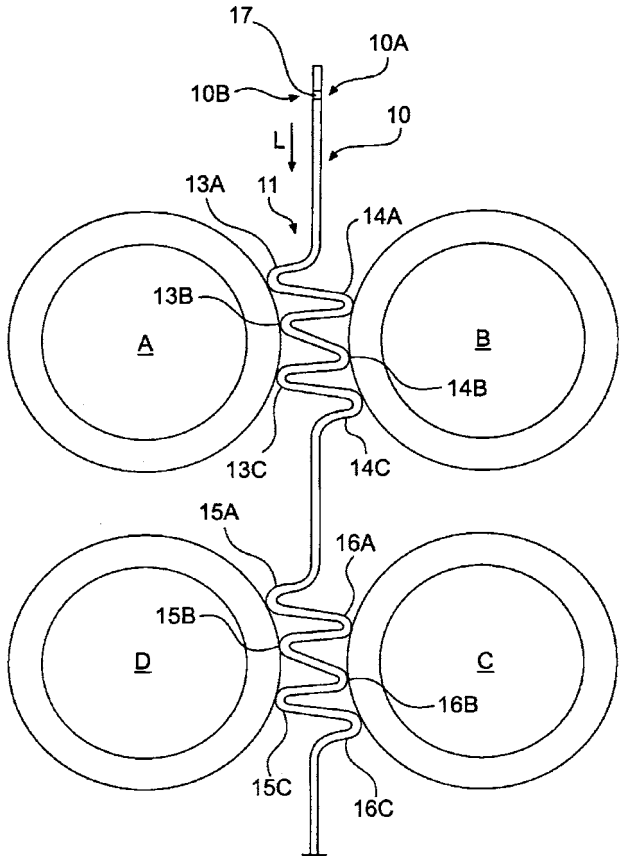
downward-traveling noise of the mud pumps. The figure shows the basic layout of the filter. A tee off of the borehole contains the mud feed pipe under pressure from a surface-mounted pump, whose outlet feeds pipe **333**. The pipe necks down to a venturi region **331**, where a Helmholtz resonator is attached on a side arm. The resonator, along with the flow restriction, acts as a filter for the pumping pressure fluctuations. A pressure transducer **341** situated downstream from this reads the telemetry signals from downstream, and additional coherent signal processing is used to extract them. Some data on the characteristic attenuations and stop bands (10–20 Hz) are given, and the patent provides a lot of background information on the drilling environment.—JAH

7,343,964

43.40.Tm ANTI-VIBRATION TUBE SUPPORT

Thomas M. Rudy *et al.*, assignors to ExxonMobil Research & Engineering Company
18 March 2008 (Class 165/162); filed 20 October 2005

This patent pertains to vibration-isolating supports for tube bundles in heat exchangers, cooling towers, steam generators, and other fluid handling equipment in which fluid flow across or within the tubes can induce vibrations. A tube support device **10**, formed from a strip of metal or other



suitable material, is provided with a series of corrugations **11** that provide resilient supports for the individual tubes **A**, **B**, **C**, and **D**.—EEU

7,347,167

43.40.Tm METHOD FOR CONTROLLING COOLING FANS

Kazuya Sugiyama *et al.*, assignors to Mazda Motor Corporation
25 March 2008 (Class 123/41.11); filed in Japan 28 February 2005

In a cooling system for an internal combustion engine that has two cooling fans, resonances are avoided by having one of the fans always operate at a rotational frequency that is higher than the engine's combustion frequency and having the other fan always operate at a rotational frequency that is lower than the combustion frequency. The fan speeds are controlled on the basis of the cooling requirements.—EEU

7,347,799

43.40.Tm ANTIVIBRATION DEVICE HAVING ROTARY FLYWEIGHTS AND AN EPICYCLIC GEARTRAIN

Thomas Manfredotti and Patrick Quesne, assignors to Eurocopter
25 March 2008 (Class 475/6); filed in France 17 February 2005

This device, suitable for reducing the vibrations in the cabin of a rotorcraft, for example, consists of two sets of two contrarotating rotors with eccentric flyweights. All rotors and unbalances are the same and operate at the same speed in parallel planes. The magnitude of the resultant force produced by the device is adjusted automatically by adjustment of the phase of one set of contrarotating rotors relative to the other.—EEU

7,355,776

43.40.Tm ACOUSTIC DAMPING MATERIAL FOR ELECTRO-OPTIC MATERIALS

Christopher P. Yakymyshyn, assignor to Tektronix, Incorporated
8 April 2008 (Class 359/245); filed 10 March 2005

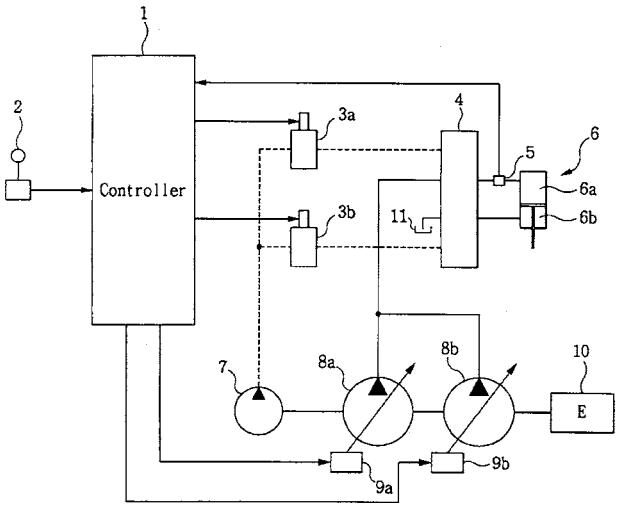
Electro-optic materials are materials whose index of refraction changes in response to electromagnetic energy that is applied to them; they are used in optical switches, optical modulators, and the like. Vibrations in such materials result from piezoelectric effects due to electromagnetic signals on electrodes connected to the electro-optic materials, causing changes in the physical dimensions of these materials and thus introducing optical noise. The patent describes an adhesive material that is combined with a ceramic crystalline material, so that the acoustical impedance of the combination is substantially equal to that of the electro-optic material. This combination is applied to the electro-optic material in order to suppress its vibrations.—EEU

7,308,789

43.40.Vn HYDRAULIC CYLINDER SUSPENSION METHOD

Jeong Kyu Lee, assignor to Volvo Construction Equipment Holding Sweden AB
18 December 2007 (Class 60/469); filed in Republic of Korea 22 March 2004

An active vibration control scheme to reduce the mechanical shock severity of starting or slapping of the boom of construction equipment is claimed. Signals from pressure sensor **5** and operator command and control



2 are analyzed by controller **1** to determine whether a sudden stop is commanded for up or down movement, while knowing the resonant period of the equipment boom. Valve controllers **3a** and **3b** are commanded to apply pressure to either the top cylinder **6a** or the bottom cylinder **6b** to counter the subsequent boom oscillation cycles.—AJC

7,311,300

43.40.Vn ACTUATOR DRIVE CONTROL DEVICE FOR ACTIVE VIBRATION ISOLATION SUPPORT SYSTEM, AND METHOD OF USING THE SAME

Hirotomi Nemoto, assignor to Honda Motor Company, Limited
25 December 2007 (Class 267/140.14); filed in Japan 10 June 2003

An active engine vibration isolation system is claimed having decreased current requirements while maintaining good isolation for the engine "shake" condition. The controller algorithm decreases the average amount of current passing though the coil.—AJC

7,344,129

43.40.Vn DAMPING METHOD AND DAMPING SYSTEM FOR HYBRID VEHICLE

Hirotomi Nemoto, assignor to Honda Motor Company, Limited
18 March 2008 (Class 267/140.14); filed in Japan 11 March 2004

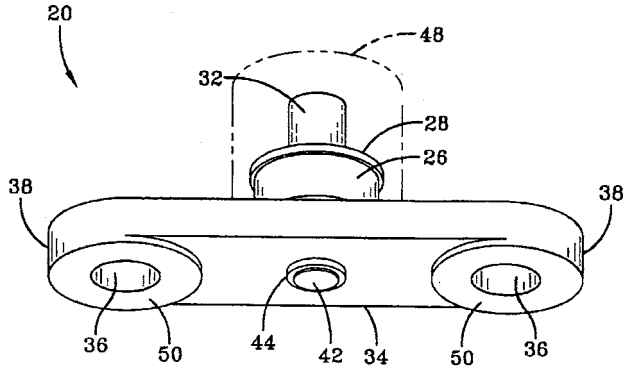
A hybrid vehicle's power unit consists of an engine and a motor (which also can act as a generator), both of which are coupled to a transmission. The unit is supported on the vehicle body frame via an active isolation system. Signals from a sensor that measures the load transmitted from the engine to the vehicle body, from a crank pulse sensor, and from an engine rotational speed sensor are fed to an electronic control unit for controlling the active isolation system and the motor. Control is accomplished in two steps: The active isolation system is actuated first, and then the motor/generator is controlled to reduce the vibrations that the isolation system does not attenuate sufficiently.—EEU

7,312,679

43.40.Yq MAGNETOSTRICTIVE STRESS WAVE SENSOR

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

This is a repeat of U.S. Patent No. 7,298,237 reviewed last month. Perhaps the claims have been changed. To reiterate: The author claims a vehicle crash shock wave sensor **20** comprising a Terfonal-D "giant magnetostrictive core" **22** next to a biasing magnet **24** and a spacer **42**, all bonded end-to-end inside tube **40** and bolted (**52** and **38**) against the vehicle frame,



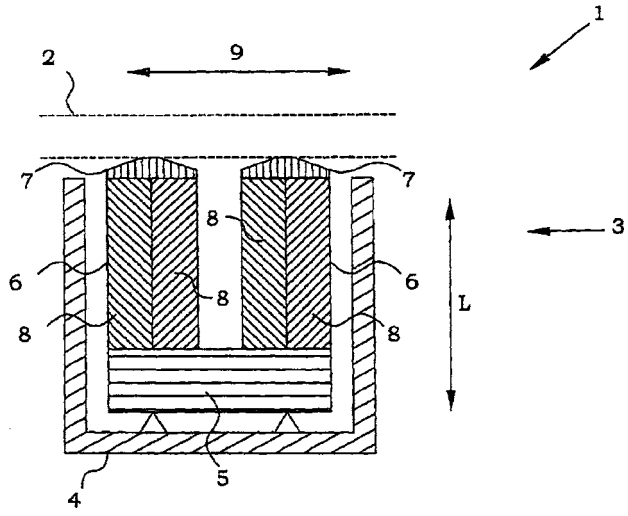
contacting at **44**. Coil **26** produces an electromotive force (EMF) from 0.2 to 2 V pulse when a vehicle shock wave occurs as the result of a crash. The assembled unit apparently has no moving parts and relies on the magnetostriction EMF excited by the structural frame vibration impulse induced by the crash.—AJC

7,355,325

43.40.Yq WIDE FREQUENCY RANGE ELECTROMECHANICAL ACTUATOR

Stefan Johansson *et al.*, assignors to Piezomotor Uppsala AB
8 April 2008 (Class 310/328); filed 15 June 2006

In this electromechanical motor the actuator **3** consists of two actuating elements **6** that are connected to a common backbone **5**, which is supported on a housing **3**. The actuating elements are piezoelectric bimorphs,



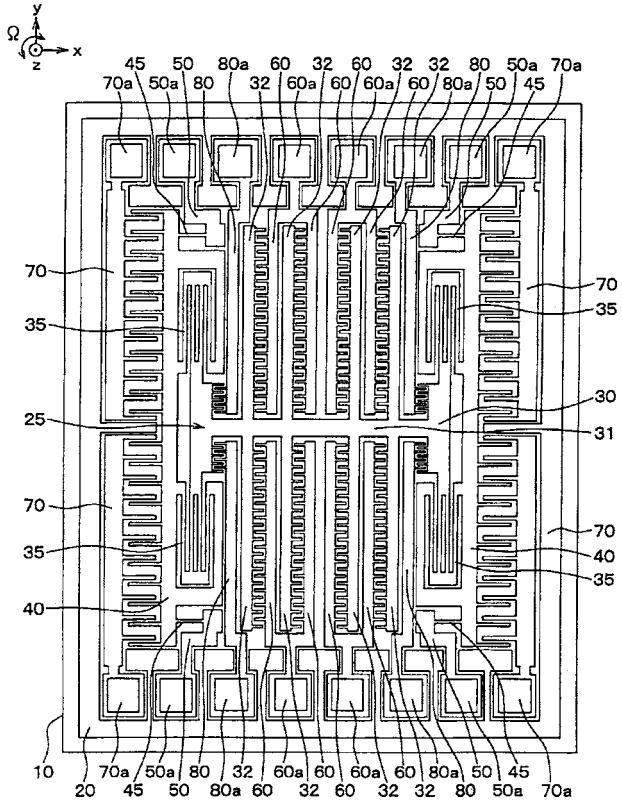
which can be made to extend longitudinally and/or to flex in response to appropriate electrical inputs. As these elements deform cyclically, their tips **7** make intermittent contact with body **2**, moving it in direction **9**.—EEU

7,318,348

43.40.Yq ANGULAR RATE SENSOR AND MOUNTING STRUCTURE OF ANGULAR RATE SENSOR

Mutsuhito Ogino, assignor to Denso Corporation
15 January 2008 (Class 73/504.14); filed in Japan 2 September 2004

This is similar to U.S. Patent No. 7,284,408 reviewed previously. An angular rate sensor **100** for vehicles is claimed using only variable capacitors. Movable electrode **25** carries vibrators **35** and **40**. Drive beam **35** and detection beam **45** are fastened to the substrate **10**. These vibrate in the *X* direction by an alternating electrostatic force due to drive voltages applied



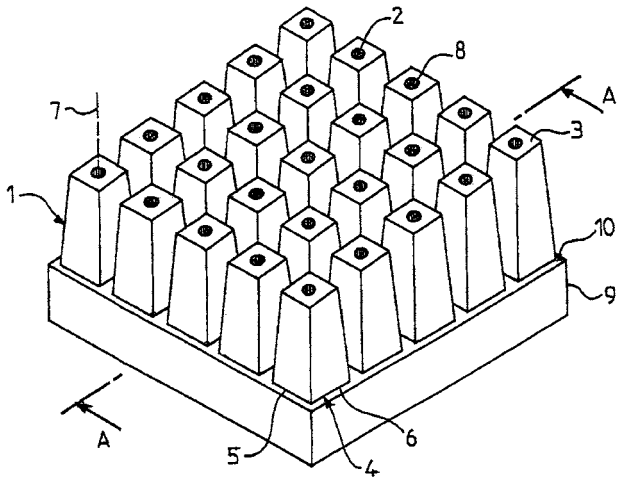
to terminals **60** and **60a**. During a turn, both positive and negative *Y* vibration results from a coriolis force on vibrating plates **32**. Coriolis (*Y*) signals are presented on terminals **50** and **50a**. Signals are received, transferred through the capacitance between fingers **65a 55a** and **65b 55b**, and are processed as $(S1+S2)-(S3+S4)$ for angular velocity and $(S2+S3)-(S1+S4)$ for acceleration. This acceleration signal is also compared to that from a separate transducer to provide an angular velocity reliability indication. Claims include dimensional limits for the beams to increase their stiffness in certain directions.—AJC

7,308,965

43.50.Gf NOISE ABATEMENT WALL

Bernard Sapoval *et al.*, assignors to Ecole Polytechnique
18 December 2007 (Class 181/210); filed in France 19 March 2002

A sound absorber for roadways is claimed where surface **1** is composed of many irregular objects **3** each of whose exposed surface area is increased over the projected surface area in a fractal sense. Sound absorption is claimed to be increased because of the surface area increase of the



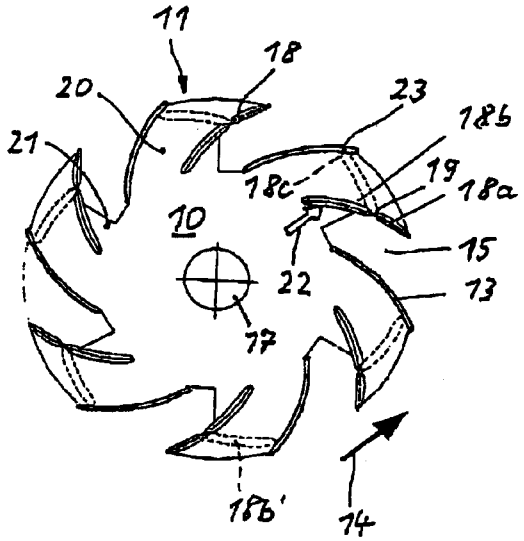
shapes in combination with an increased sound amplitude between and near these shapes.—AJC

7,309,935

43.50.Gf FAN IMPELLER FOR ELECTRICAL MACHINES

Horst Braun and Hans-Joachim Lutz, assignors to Robert Bosch GmbH
18 December 2007 (Class 310/62); filed in Germany 9 November 2001

Vehicle alternators need the most cooling air flow at low RPM. At high vehicle speeds, fan noise becomes excessive. A quieter alternator cooling



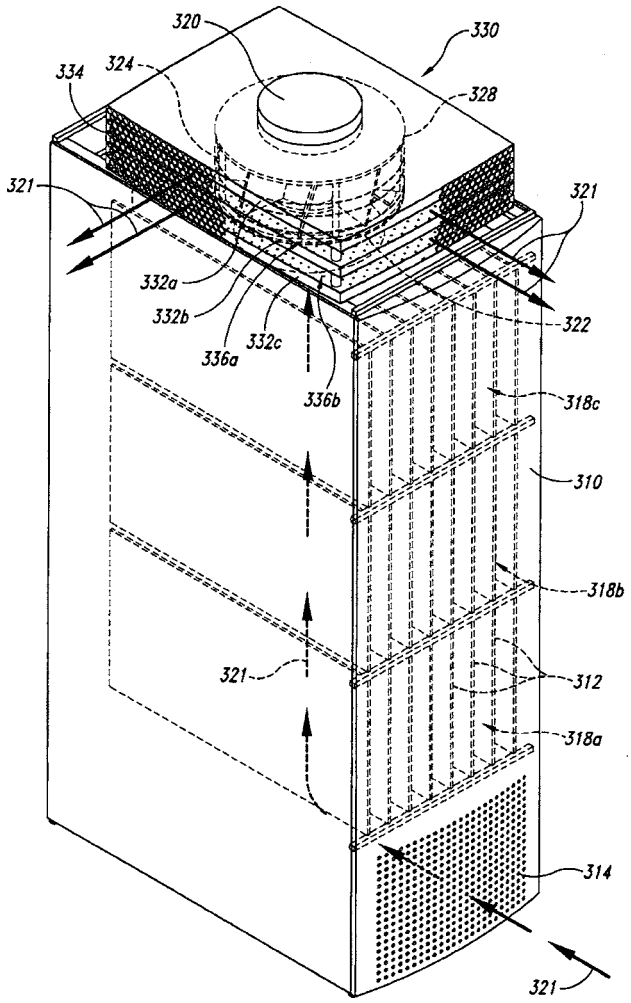
fan, having movable vanes **18b** that swivel out to position **c** at higher RPM values is claimed, thereby reducing air flow and noise.—AJC

7,314,113

43.50.Gf ACOUSTIC ABSORBERS FOR USE WITH COMPUTER CABINET FANS AND OTHER COOLING SYSTEMS

Wade J. Doll, assignor to Cray Incorporated
1 January 2008 (Class 181/225); filed 14 September 2004

A quiet supercomputer **318–321** cooling fan **330** requiring minimal floor space is claimed where centrifugal fan wheel **328** is surrounded by acoustical filter **322** through which exhaust air **321** is discharged. Filter **22**



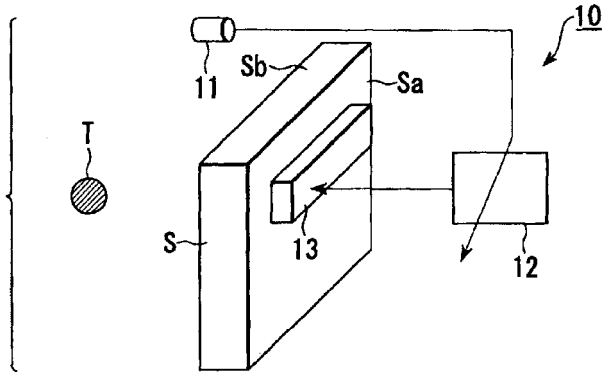
comprises sound absorbing panels **332** aligned parallel to air flow **321** terminating discharge openings **336** that are arranged in all directions around fan system **330**.—AJC

7,352,870

43.50.Ki ACTIVE SOUND MUFFLER AND ACTIVE SOUND MUFFLING METHOD

Akihiko Enamito *et al.*, assignors to Kabushiki Kaisha Toshiba
1 April 2008 (Class 381/71.2); filed in Japan 29 March 2002

Here is an active noise cancellation system that specifically addresses low frequency sound that would be diffracted over the top of an outdoor



barrier, e.g., along a highway. In this manner, by adding noise cancellation, you can reduce the necessary height of the barrier.—CJR

7,352,874

43.50.Ki APPARATUS FOR ACOUSTICALLY IMPROVING AN ENVIRONMENT AND RELATED METHOD

Andreas Raptopolous, London SE1 1GJ, United Kingdom and Michael Kieslinger, 100 Vienna, Austria
1 April 2008 (Class 381/152); filed in United Kingdom 16 November 1999

This electronic sound screening system transforms noise (a sow's ear) into a pleasant nondisturbing sound environment (a silk purse). The curtain or wall can be seen as a smart textile, with a passive and an active element. The passive element absorbs sound, and reduces the noise level by a few decibels. The active element transforms the remaining sound, and gives it a pleasant quality. This transformation is done with signal processing, and the resulting pleasant sound is played back through loudspeakers connected to the curtain or wall.—CJR

7,344,772

43.55.Rg SOUND-DAMPING PROFILED MEMBER

Marc Rehfeld *et al.*, assignors to Saint-Gobain Glass France
18 March 2008 (Class 428/304.4); filed in France 31 July 2002

The patent describes a plastic-based damping strip that is to be inserted between two elements, like the layers of glass in an automobile. This results in improved sound reduction.—CJR

7,343,715

43.55.Ti SOUND-PROOF WALL MADE OF FRP, AND METHOD OF PRODUCING THE SAME

Toshihiro Ito *et al.*, assignors to Toray Industries, Incorporated
18 March 2008 (Class 52/309.16); filed in Japan 17 May 2001

This panel is part of a sound barrier wall assembly using fiberglass-reinforced plastic. The panel is light-weight and durable outdoors.—CJR

7,353,638

43.55.Ti SOUND RESISTANT ENCLOSURE

Michael Salemi, assignor to Salemi, Incorporated
8 April 2008 (Class 52/27.5); filed 10 February 2006

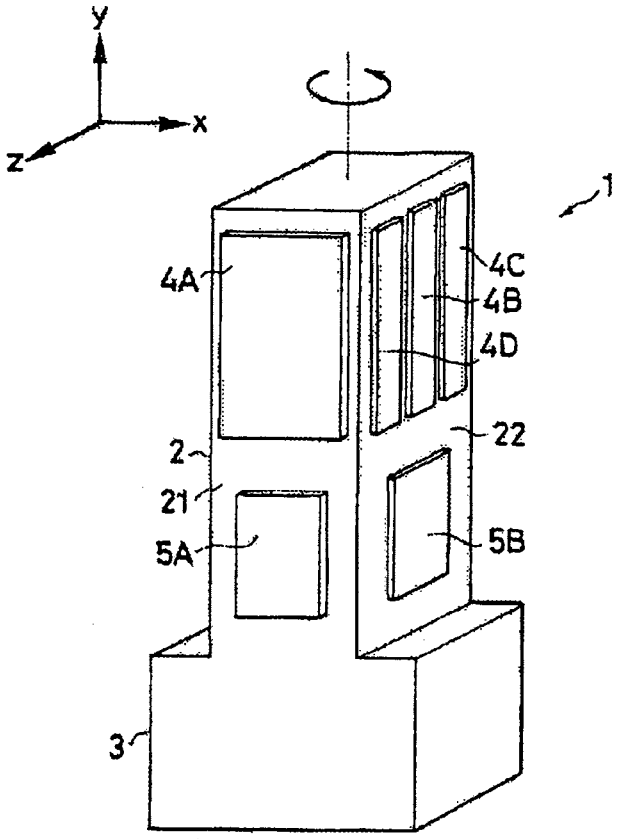
This enclosure encapsulates its occupant and blocks environmental noise so that the occupant can make cellular phone calls without hearing intrusive distractions. It seems very much like what used to be referred to as a phone booth.—CJR

7,348,716

43.58.Wc PIEZOELECTRIC GYRO ELEMENT AND PIEZOELECTRIC GYROSCOPE

Seiichiro Ogura and Yoshihiro Kobayashi, assignors to Seiko Epson Corporation
25 March 2008 (Class 310/370); filed in Japan 16 December 2004

This patent discloses the design of a tuning-fork style gyro with multiple pickoff and excitation electrodes. The figure shows electrodes **4A**, **B**, **C**, which are used to excite a circular precession of the piezoelectric crystal element **2**. The motion is picked off by electrodes **5A**, **B** to provide sensing



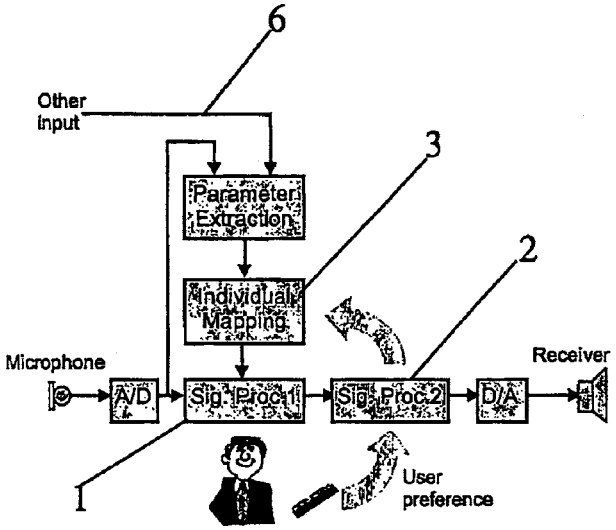
of the motion and phase of the precession. Although not a particularly novel construction for the resonant element **2**, the electrode configuration is well thought out, and the principles are applicable to a large class of mechanically resonant gyros.—JAH

7,340,231

43.66.Ts METHOD OF PROGRAMMING A COMMUNICATION DEVICE AND A PROGRAMMABLE COMMUNICATION DEVICE

Thomas Behrens *et al.*, assignors to Oticon A/S
4 March 2008 (Class 455/173.1); filed in Denmark 5 October 2001

The hearing aid wearer adjusts signal processing parameters of the hearing aid for best sound quality in a training session while listening to his or her own voice through the device. These preferred parameters are stored and are used automatically whenever the user's own voice is detected.—DAP

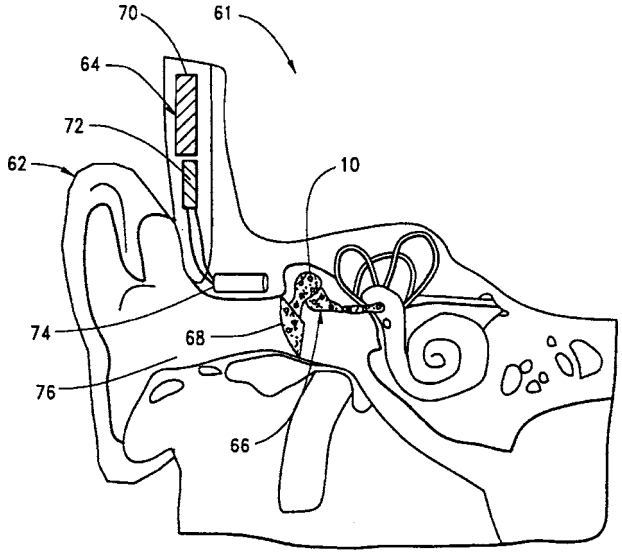


7,344,491

43.66.Ts METHOD AND APPARATUS FOR IMPROVING HEARING

Charles E. Seeney and Kenneth J. Dormer, assignors to Nanobiomagnetics, Incorporated
18 March 2008 (Class 600/25); filed 14 October 2004

Magnetically responsive nanospheres are moved into cells on the tympanic membrane or on the ossicular chain of the middle ear of a mammal using a first magnetic field. Thereafter, an oscillating magnetic field, produced by an electromagnetic coil in response to incoming sound waves, is used to vibrate the magnetically responsive nanoparticles contained in the supported nanospheres.—DAP

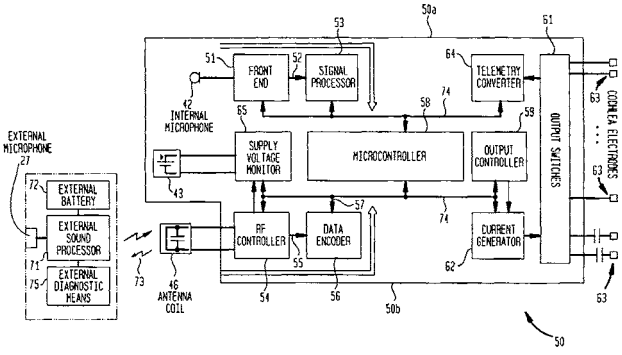


7,346,397

43.66.Ts COCHLEAR IMPLANT

David Money *et al.*, assignors to Cochlear Limited
18 March 2008 (Class 607/57); filed in Australia 30 June 2000

When there is insufficient charge remaining in a rechargeable battery sealed into a cochlear implant housing, a warning signal is generated and recognized by the implantee. Thereafter, the electrical stimulation for a cochlear implant can be deactivated in response to the detection of a predetermined acoustic signal. The acoustic signal may be noise generated by



tapping on the implantee's head. The system utilizes a wire antenna coil to receive power for inductive charging from an external source and also function as part of a bidirectional rf communication link to an external controller.—DAP

7,349,549

43.66.Ts METHOD TO LOG DATA IN A HEARING DEVICE AS WELL AS A HEARING DEVICE

Herbert Bächler *et al.*, assignors to Phonak AG
25 March 2008 (Class 381/314); filed 27 January 2004

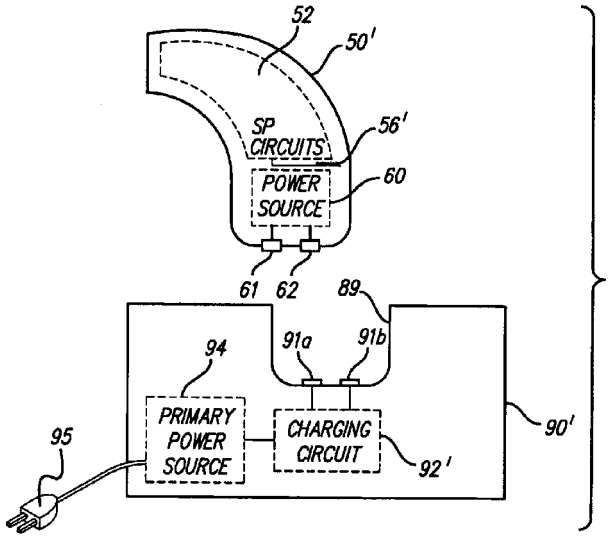
A hearing aid uses neural networks to self-learn. User preferences are logged with statistical environmental data using an external device. The hearing aid fitter is provided with stored diagnostic information about previous fitting history and adjustments. Current time information from the external source is transmitted to synchronize the hearing aid. Acoustic signals and synthetic speech information are generated in the hearing aid for the user based on previously set time or date information. Motivations include triggering data logging times and announcing appointments and times to take medicine.—DAP

7,349,741

43.66.Ts COCHLEAR IMPLANT SOUND PROCESSOR WITH PERMANENTLY INTEGRATED REPLENISHABLE POWER SOURCE

Albert A. Maltan *et al.*, assignors to Advanced Bionics, LLC
25 March 2008 (Class 607/57); filed 30 September 2003

A rechargeable battery is sealed with two coils and a magnetic sensor into the housing of a cochlear implant sound processor. The first coil receives power inductively and the processor receives control signals electromagnetically when the sensor indicates that they are in close proximity to an



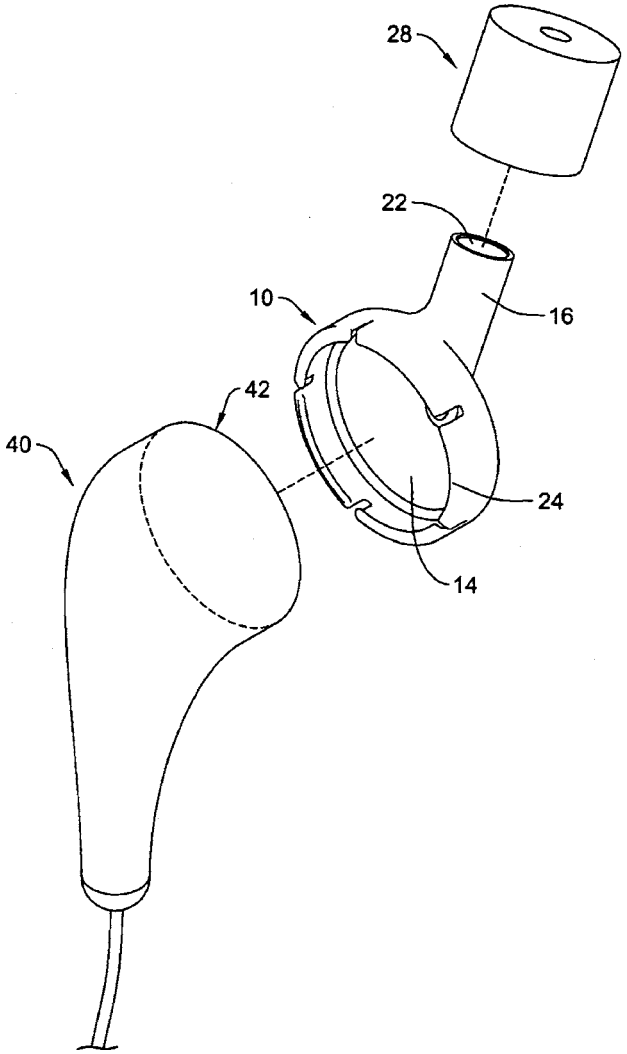
external charging source base station. The sound processor section, which may also fit into the base station for hard-wired recharging, transmits status signals back to the base station via the first antenna-coil and provides power and electrical stimulation signals for the implantable portion via the second coil.—DAP

7,349,550

43.66.Ts EARBUD ADAPTER

Robert J. Oliveira *et al.*, assignors to Hearing Components, Incorporated
25 March 2008 (Class 381/380); filed 7 January 2004

An adapter has a first side with a projection on which a resilient foam sleeve configured to fit into the wearer's ear canal is mounted, and an opposite second side that has a releasable attachment mechanism for connec-



tion to an earbud or earbud-type headphone. The sleeve seals the ear canal to help prevent the wearer from receiving extraneous sounds.—DAP

7,352,871

43.66.Ts APPARATUS FOR COMMUNICATION AND RECONNAISSANCE COUPLED WITH PROTECTION OF THE AUDITORY SYSTEM

Ben T. Mozo, Enterprise, Alabama
1 April 2008 (Class 381/72); filed 24 July 2003

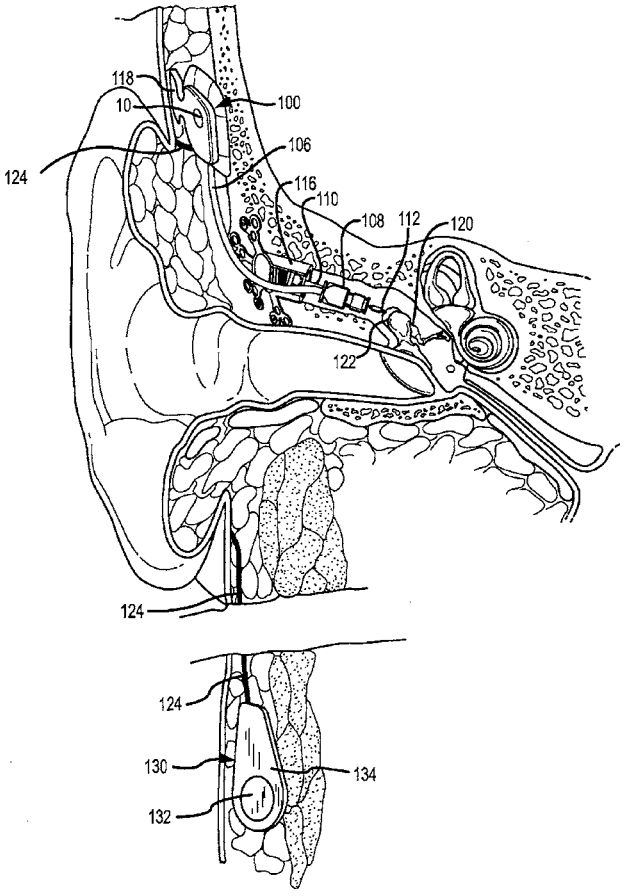
This binaural device protects hearing by incorporating either an insert or circumaural hearing protector into which a microphone and earphone are incorporated. Input to the earphone is through either the microphone or radio circuit. Output of speech to the ear of the wearer and to a radio is through the microphone circuit. Adjustable gain to the ear is provided to optimize signal-to-noise ratio and is limited by the battery voltage.—JE

7,354,394

43.66.Ts SOFT TISSUE PLACEMENT OF IMPLANTABLE MICROPHONE

William Howard Slattery III *et al.*, assignors to Otologics, LLC
8 April 2008 (Class 600/25); filed 20 June 2006

Pickup by an implantable middle ear hearing system of undesirable, skull-borne vibrations is reduced, including those produced from talking, chewing, and the implantable transducer. The microphone is isolated mechanically to reduce these vibrations by mounting it at a subcutaneous



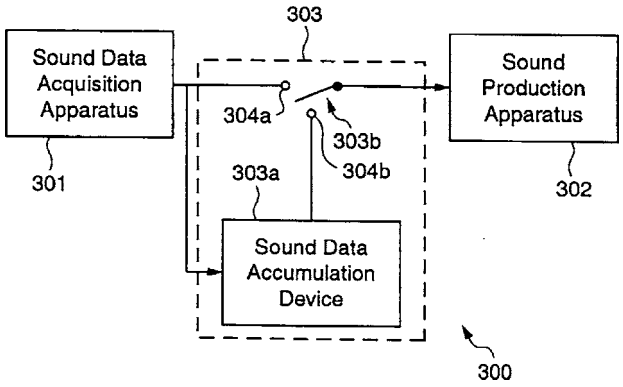
location on soft tissue, such as the patient's neck or pinna, spaced at least 2–5 mm from any underlying bone. The microphone output is routed via signal wire subcutaneously to the implanted auditory stimulation device.—DAP

7,356,153

43.66.Ts HEARING ASSISTIVE APPARATUS HAVING SOUND REPLAY CAPABILITY

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

Sounds in the vicinity of the hearing device wearer are acquired and stored. Predetermined time segments of these sounds are later played back through a sound production apparatus mounted at or near the wearer's ear.



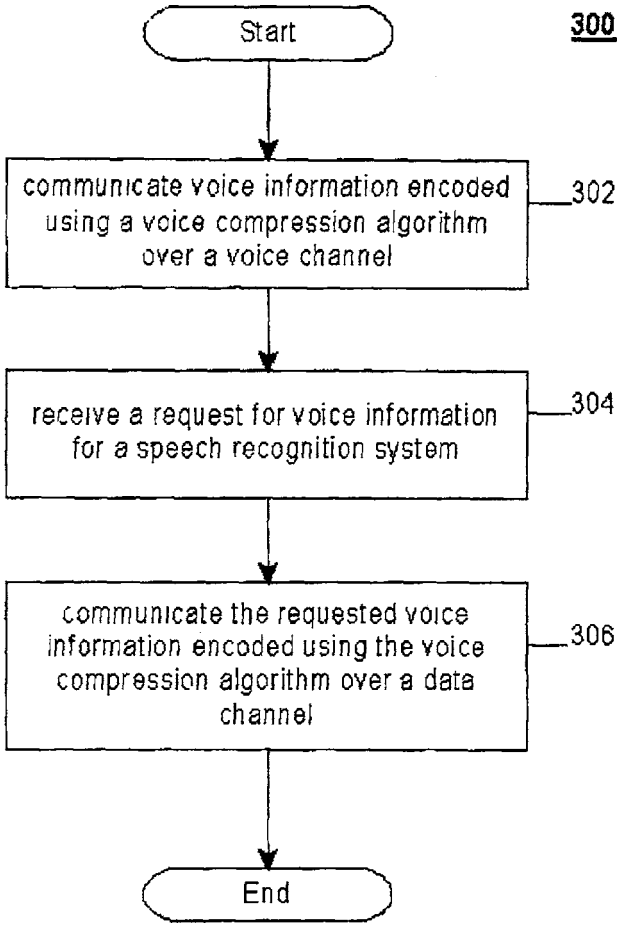
Also available under wearer control are a mute mode and a non-playback/real time mode. The sound acquisition and replay portions of the system may be physically separated from the sound production portion using a wireless link for communication.—DAP

7,346,496

43.72.Ne METHOD AND APPARATUS TO PERFORM SPEECH RECOGNITION OVER A DATA CHANNEL

Michael Stanford, assignor to Intel Corporation
18 March 2008 (Class 704/201); filed 8 January 2007

The goal of this patent is to improve speech recognition accuracy over wireless communication systems that have limited bandwidth for voice transmissions. Better speech quality is produced using wider bandwidth data



transmission channels instead of voice transmission channels, which should improve speech recognition. To facilitate communication over data channels, voice information is digitized and encoded using a compression algorithm.—DAP

7,349,844

43.72.Ne MINIMIZING RESOURCE CONSUMPTION FOR SPEECH RECOGNITION PROCESSING WITH DUAL ACCESS BUFFERING

Dieter Staiger, assignor to International Business Machines Corporation
25 March 2008 (Class 704/231); filed in the European Patent Office 14 March 2001

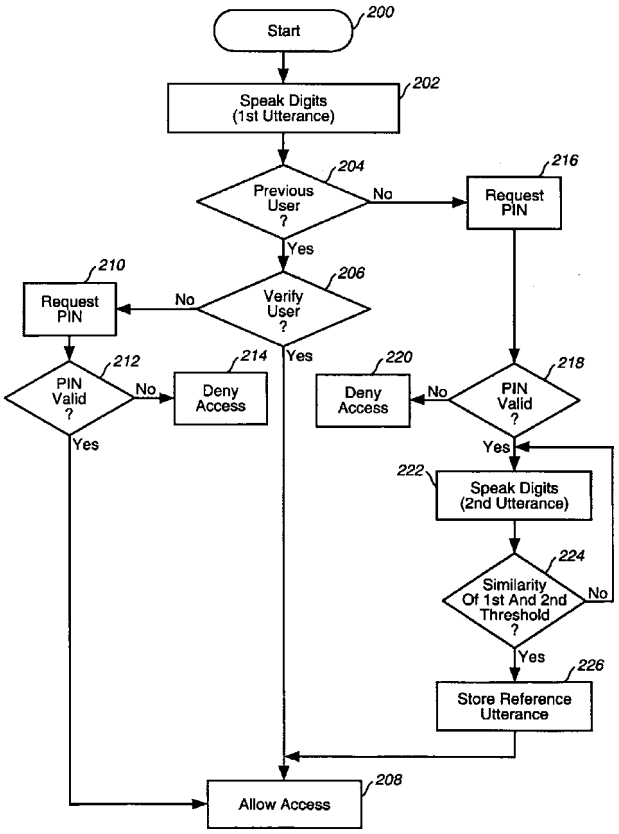
Sequential operation is used to lower power consumption during speech recognition. The core processor is invoked only after the start of preprocessing and is utilized only when the amount of data in a dual access storage buffer goes above a predetermined threshold level. Likewise, back-end processing is invoked only after core processing has started and sufficient data are available in the dual-access storage buffer.—DAP

7,356,134

43.72.Ne METHOD OF ACCESSING A DIAL-UP SERVICE

Robert Wesley Bossemeyer, Jr., assignor to SBC Properties, L.P.
8 April 2008 (Class 379/93.03); filed 3 December 2004

When a user is dialing a service number, speaker-independent speaker recognition is used to verify a previous user by recognizing digits in a first utterance spoken. If the user cannot be verified, the system requests the user



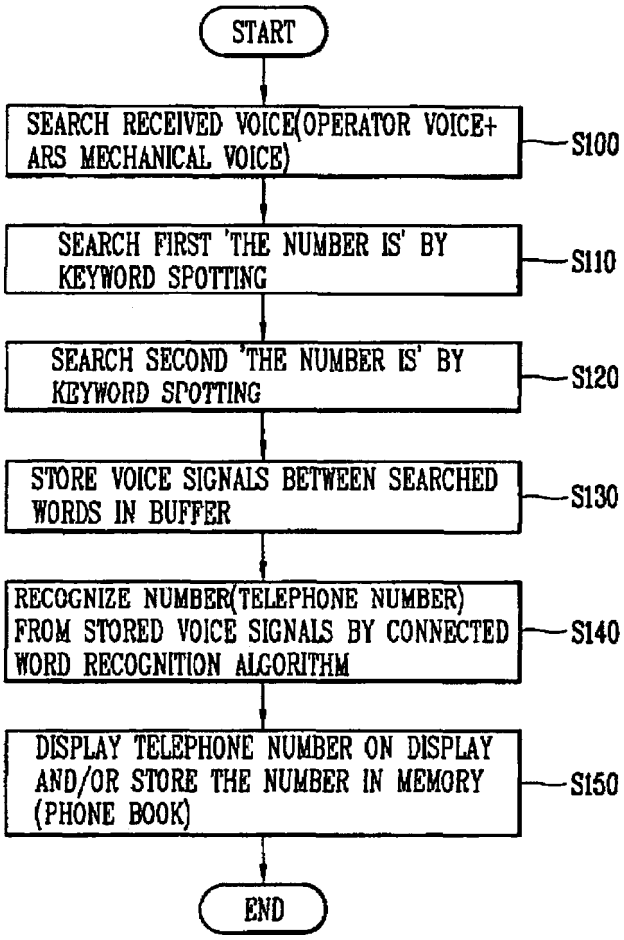
to speak or enter a personal identification number (PIN). If the PIN is valid, before access is granted, the user may be requested to speak a second utterance that is compared for similarity to the first utterance.—DAP

7,356,356

43.72.Ne TELEPHONE NUMBER RETRIEVAL SYSTEM AND METHOD

Chan-Woo Kim *et al.*, assignor to LG Electronics Incorporated
8 April 2008 (Class 455/563); filed in Republic of Korea 29 October 2003

Motivations for this patent are (1) to improve the accuracy of telephone number recognition in noisy conditions that are provided by the audio output of an automatic response service and (2) to provide the ability to display and store the telephone numbers on a mobile communication device. Designated words or phrases are added per a downloaded reference vector



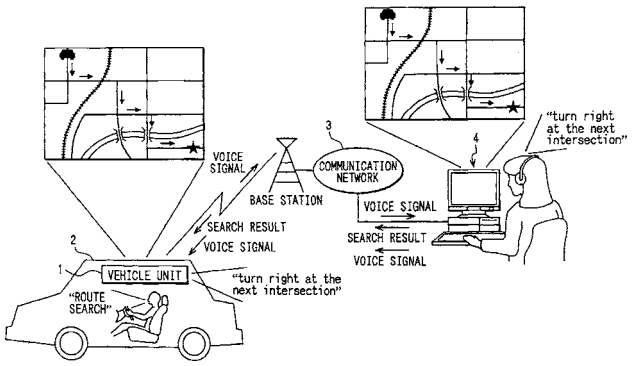
sequence before and after digits in the telephone number voice output from the automatic response service. These words are recognized by a keyword spotting routine. The numbers connecting the designated words are extracted and processed to create a recognized telephone number.—DAP

7,356,471

43.72.Ne ADJUSTING SOUND CHARACTERISTIC OF A COMMUNICATION NETWORK USING TEST SIGNAL PRIOR TO PROVIDING COMMUNICATION TO SPEECH RECOGNITION SERVER

Toshiyuki Ito and Hiroshige Asada, assignors to DENSO Corporation
8 April 2008 (Class 704/270.1); filed in Japan 25 June 2002

Goals for this patent are increased reliability of voice recognition and reduced computing and storage requirements for the voice recognition server used in vehicular location and routing applications. The sound characteristics of a communication network, such as a cellular telephone net



work, are adjusted for improved recognition accuracy. A voice test pattern signal, transmitted by the voice recognition server to the communication terminal in the vehicle, is looped back and used for the adjustments.—DAP

7,353,056

43.80.Vj OPTIMIZED SWITCHING CONFIGURATIONS FOR RECONFIGURABLE ARRAYS OF SENSOR ELEMENTS

Christopher Robert Hazard *et al.*, assignors to General Electric Company
1 April 2008 (Class 600/407); filed 29 October 2004

Groups of elements in a two-dimensional transducer array are connected dynamically by a switching network to match the shape of an ultrasonic wavefront.—RCW

7,347,821

43.80.Vj ADAPTIVE PROCESSING OF CONTRAST ENHANCED ULTRASONIC DIAGNOSTIC IMAGES

Daniel Skyba and Damien Dolimier, assignors to Koninklijke Philips Electronics N.V.
25 March 2008 (Class 600/443); filed 21 June 2004

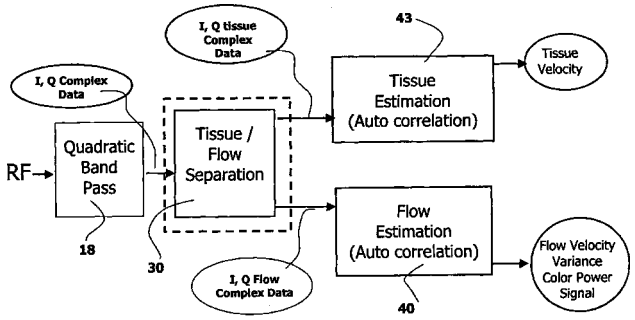
Endocardial and epicardial tissue borders are delineated during ultrasound imaging of the heart in the presence of a contrast agent by defining a set of threshold values associated with imaging depth and using the values in an adaptive processing algorithm.—RCW

7,347,820

43.80.Vj PHASED ARRAY ACOUSTIC SYSTEM FOR 3D IMAGING OF MOVING PARTS

Odile Bonnefous, assignor to Koninklijke Philips Electronics N.V.
25 March 2008 (Class 600/437); filed in the European Patent Office 6 November 2002

This system forms multiple receiver beams from a single transmit beam and uses the spatial and temporal correlation properties of signals in



the received beams to separate and to image echoes from tissue and blood flow.—RCW

7,350,522

43.80.Vj SCANNING METHOD FOR APPLYING ULTRASONIC ACOUSTIC DATA TO THE HUMAN NEURAL CORTEX

Thomas Patrick Dawson, assignor to Sony Corporation
1 April 2008 (Class 128/897); filed 12 April 2004

A desired sensory perception in the neural cortex is produced by a scanned acoustic signal. The initial scanning of the signal is in a predetermined pattern. The pattern is then modified to fill spaces in the pattern until the entire cortical region of interest is covered.—RCW

7,354,400

43.80.Vj ULTRASONOGRAPHIC METHOD AND ULTRASONOGRAPHIC DEVICE

Katsunori Asafusa and Hiroshi Kanda, assignors to Hitachi Medical Corporation
8 April 2008 (Class 600/437); filed in Japan 7 January 2003

The signal in an ultrasonic beam is encoded using a set of modulation codes. A corresponding set of codes is used for demodulation of signals in the ultrasonic receiver beam. The transmission–reception step is repeated with the arrangement of the modulation code reversed. Decoded signals from the two sets of codes are combined to form an image.—RCW

LETTERS TO THE EDITOR

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

Comment on "Ear Asymmetries in middle-ear, cochlear, and brainstem responses in human infants" [J. Acoust. Soc. Am. 123, 1504–1512] (L)

Yvonne Sininger, Ph. D., Professor, Head & Neck Surgery
University of California Los Angeles, 62-132 Center for the Health Sciences, Box 951624, Los Angeles, California 90095-1624

Barbara Cone, Ph.D., Professor, Speech, Language and Hearing Sciences[a)]
University of Arizona, P.O. Box 210071, Tucson, Arizona 85721

(Received 1 April 2008; revised 12 June 2008; accepted 12 June 2008)

Following Sininger and Cone-Wesson [Science **305**, 1581], Sininger and Cone-Wesson [Hear. Res. **212**, 203–211], Keefe *et al.* [J. Acoust. Soc. Am. **123**(3), 1504–1512] described ear asymmetries in middle ear, cochlear, and brainstem responses of infants. Keefe *et al.* state that their data do not support the findings of Sininger and Cone-Wesson [Science **305**, 1581] who found asymmetries in evoked otoacoustic emissions and auditory brainstem responses and proposed that stimulus-directed asymmetries in processing may facilitate development of hemispheric specialization. The Keefe *et al.* findings, in fact, replicated and extended the findings of Sininger and Cone-Wesson (2004, 2006) and support, rather than refute, the conclusions. Keefe *et al.* controlled neither the background noise nor averaging time across test conditions (ear or otoacoustic emission type) and thus their separate analyses of signal and noise magnitude exceed the limitations imposed by the data collection methods.
 [DOI: 10.1121/1.2956481]

PACS number(s): 43.64.Jb [BLM] Pages: 1401–1403

Sininger and Cone-Wesson (2004) were the first to suggest that stimulus-directed laterality was evident in evoked otoacoustic emissions (EAOEs) obtained from human newborns. Using data collected to evaluate neonatal hearing screening protocols, they demonstrated that the signal-to-noise ratios (SNRs) of distortion-product otoacoustic emissions (DPOAEs) were larger for the right versus left ear, and transient-evoked OAEs (TEOAEs) SNRs were larger for the left versus right ear. Keefe *et al.* (2008) have replicated and extended these findings, as well as the work of Sininger and Cone-Wesson (2006) who discerned right ear laterality, evidenced by shorter latencies, for the click-evoked auditory brainstem response (ABR). Keefe *et al.* extended these findings by considering whether middle ear transmission characteristics, as measured using reflectance, could account for this laterality. Their data indicate that middle-ear factors cannot account for the EOAE and ABR asymmetries described by Sininger and Cone-Wesson. We appreciate this important contribution but argue that their analyses of the infant EOAE data were flawed.

Figure 1 of Keefe *et al.* shows a comparison of the SNR of the DP- and TEOAEs by ear in the two studies. Clearly, the analyses of Keefe *et al.* demonstrate exactly the same findings as those of Sininger and Cone-Wesson (2004). Yet, they present an argument that their data *do not* support the ear asymmetries for EOAEs.

Both Sininger and Cone-Wesson (2004, 2006) and Keefe *et al.* used preexisting data sets to evaluate asymmetry. These data sets were obtained to evaluate newborn hearing screening technologies, with data collection methods appropriate to that purpose. Screening technologies will halt data collection as soon as a response can be reliably detected, in this case, as soon as the SNR reached a criterion point. For this reason, background noise was not controlled in the recordings. Consequently, it is impossible to say how OAE amplitude and noise are independently contributing to the SNR in this data set. Keefe *et al.* dismiss issues regarding the limitations of the stopping rules applied in the original data collection, although they acknowledge this in the discussion. We argue that given the limitations of the way in which the data were acquired, the only appropriate description of these data across ears is the SNR at the time the test was stopped.

Independent analyses of OAE signal or noise should *not*

[a)]Author to whom correspondence should be addressed; electronic mail: conewess@u.arizona.edu

TABLE I. Mean OAE noise level in dB SPL at the conclusion of averaging. Data are from Sininger and Cone Wesson (2004).

	DPOAE (1.5-4k)	TEOAE
Right ear	1.0560	7.91
Left ear	0.3539	7.00

be performed on this particular data set. Backus (2007) discusses the error of estimating OAE magnitude without taking into consideration the noise levels. He points out that noise contamination on magnitude estimation is least at high SNRs. However, the SNR of the data under discussion was only 3 dB or better leaving the potential contamination of the magnitude by noise as a distinct question. Keefe *et al.* used SNR criteria ranging from 2 to 6 dB for data inclusion, neither of which affected the outcome of the analyses. These criteria were too lenient to address the issue of noise contamination of the signal strength. If the extracted signals (response magnitude) are to be compared across conditions, in this case, for right versus left ears, the noise contribution to each recording should have been controlled, but was not. It is important to control for the number of sweeps or amount of noise when the background noise of an averaged response (in this case, OAE) is to be compared across ears because noise is reduced exponentially by averaging (Don and Elberling, 1994). Likewise, it is invalid to make comparisons of signal magnitude across ears without control of noise. Test stopping rules were based upon SNR, not upon a criterion number of sweeps nor a "low noise" criterion. Noise was greater in the right ear when the test was stopped. The estimate of response magnitude is contaminated by that noise (Backus, 2007).

In fact, we have now determined that the number of sweeps or duration of averaging time differed significantly across ears for the data reported in Science (2004). The right ear TEOAEs, which demonstrated higher SNR than those of the left ear, reached the stopping rule criterion in fewer averaged sweeps than the left ear. It took an average of 113 sweeps to reach criterion SNR for the right ear but 128 sweeps for the left ($t=3.73$, $P=0.0002$). For DPOAEs, the duration of averaging time rather than the number of sweeps was calculated. The analysis of averaging time is consistent with the finding of larger DPOAE SNRs for the left compared to right ear; the SNR criterion was met after an average of 100 s for the left ear while it took an average of 106 s for the right ear ($t=2.94$, $P=0.0033$). It can be argued that if the number of sweeps or averaging time had been controlled, the ear asymmetry in SNR amplitude, and even for signal (response) level would have been accentuated.

It is interesting that the analyses of Keefe *et al.* show that noise is always higher in the right ear, for both DP- and TEOAEs, when the SNR stopping rule criteria were met. Re-analysis of the data used by Sininger and Cone-Wesson (2004) also showed this trend (see Table I). This means that signal levels had to be higher in the right ear, to meet the criterion SNR (and all infants included in the analyses met the SNR criteria for both ears). For DPOAE, noise is higher in the right ear and the subsequent SNR obtained for the left ear is larger than that in the right (refer to Keefe *et al.* Fig. 1), as it was in Sininger and Cone-Wesson (2004). The Keefe *et al.* TEOAE left vs. right difference did not meet their *a priori* statistical criterion, although it is in the same direction as the Sininger and Cone-Wesson data.[1] The fact that there is no right ear advantage for TEOAE SNR is surprising, given the preponderance of OAE studies (cited in the introduction of Keefe *et al.*) that show that spontaneous and evoked OAEs are more prevalent and larger in the right ear compared to the left and the recently published data from 30 000 infants (Berninger, 2007) that reconfirms the right ear TEOAE SNR advantage. The failure to find a right ear advantage for TEOAE SNR, although they do so for the TEOAE signal magnitude, again highlights the problem of estimating signal and noise separately in this data set.

Keefe *et al.* do not acknowledge current theory that there is not an isolated "right ear advantage" but rather, at the level of the auditory cortex, as well as measured performance by ear, that the type of stimulus guides the lateralization. Processing of broadband, rapidly changing or temporally complex auditory signals including the fine structure of speech is facilitated in the auditory areas of the left hemisphere and processing of slowly changing narrow band or tonal stimuli including tones and music is primarily accomplished in the auditory areas of the right temporal lobe (Zatorre and Belin, 2001). It has long been known that the ear contralateral to each hemisphere may show similar functional superiority based on the task. Specifically, the right ear has an "advantage" in a dichotic listening environment for the processing of speech sounds, particularly consonant perception, but the left ear appears to have preference for the processing of tonal sounds and music (Kimura, 1961; Kimura, 1964).

Although Keefe *et al.* stated that their results do not support the theory that there is a "*stimulus driven, peripherally based factor…that might influence the functional maturation of hemispheric specialization in the brain*," in fact, they replicated the Sininger and Cone-Wesson EOAE and ABR results. Sininger and Cone-Wesson suggested that differences observed in EOAE SNRs may have been due to asymmetry in the function of the medial olivo-cochlear system (Guinan, 2006). Keefe *et al.* have not provided data that support an alternative model or mechanism nor have they tested such a model or mechanism. In summary, we believe the data of Keefe *et al.* replicate and extend our previous work and can be interpreted to support the theory of stimulus-directed laterality of ear and brain function. We agree that there are many unanswered questions regarding asymmetry of the OAE based on stimulus type that are worthy of consideration, as Keefe *et al.* suggest. New experiments must be conducted to address the questions of laterality and its mechanisms on OAE magnitude, noise, and SNR across ears and by stimulus type.

[1]Their SNR for TEOAEs indicates a right ear advantage, but their study may have been underpowered (compared to Sininger and Cone-Wesson) to find a statistically significant effect.

Backus, B. C. (**2007**). "Bias due to noise in otoacoustic emission measurements," J. Acoust. Soc. Am. **121**, 1588–1603.
Berninger, E. (**2007**). "Characteristics of normal newborn transient-evoked otoacoustic emissions: Ear asymmetries and sex effects," Int. J. Audiol.

46, 661–669.
Don, M., and Elberling, C. (**1994**). "Evaluating residual background noise in human auditory brain-stem responses," J. Acoust. Soc. Am. **96**, 2746–2757.
Keefe, D. H., Gorga, M. P., Jesteadt, W., and Smith, L. M. (**1961**). "Ear asymmetries in middle-ear cochlear and brainstem responses in human infants," J. Acoust. Soc. Am. **123**, 1504–1512.
Kimura, D. (**1961**). "Cerebral dominance and the perception of verbal stimuli," Can. J. Psychol. **15**, 166–171.
Kimura, D. (**1964**). "Left-ear differences in the perception of melodies Quarterly Journal," Q. J. Exp. Psychol. **16**, 355–358.
Sininger, Y. S., and Cone-Wesson, B. (**2004**). "Asymmetric cochlear processing mimics hemispheric specialization," Science **305**, 1581.
Sininger, Y. S., and Cone-Wesson, B. (**2006**). "Lateral asymetry in the ABR of neonates: Evidence and mechanisms," Hear. Res. **212**, 203–211.
Zatorre, R. J., and Belin, P. (**2001**). "Spectral and temporal processing in human auditory cortex," Cereb. Cortex **11**, 946–953.

Radiation force in nonlinear, focused beams (L)

L. A. Ostrovsky[a)]
Zel Technologies/NOAA Earth System Research Laboratory/University of Colorado, 325 Broadway, Boulder, Colorado 80027

(Received 18 February 2008; accepted 11 June 2008)

The effect of cumulative nonlinear distortions in a focused ultrasonic wave on the dissipative radiation force (RF) is analyzed using a simplified analytical model separating the stages of geometrical focusing and the subsequent diffraction in the focal area. It is shown that in the absence of shocks, nonlinear steepening can significantly amplify the RF as compared with a harmonic wave. Two different types of dissipation, one corresponding to a classical viscous fluid and another typical of many biological tissues, are considered.
 [DOI: 10.1121/1.2956473]

PACS number(s): 43.25.Qp, 43.80.Gx, 43.25.Cb, 43.35.Wa [RMW] Pages: 1404–1407

I. INTRODUCTION

Acoustic radiation force (RF), is, in general, defined as a period-average force exerted on a medium by a sound wave. The RF concept has a long history. At first the main attention was drawn to the average sound action on small bodies and interfaces. Later, however, distributed forces (similar to electromagnetic "ponderomotive" force) became of no less interest. At present, there exists a large body of literature concerning "radiation force," "radiation pressure," and "radiation stress" in fluids and solids. For a more detailed outline of the problem in the context of this paper see, e.g., Ref. 1 and references therein. One of the earlier and most through studies was carried out by Eckart,[2] who derived equations for the average force and the resulting motion in a quasi-harmonic wave in fluids. As expected, in fluids the RF-induced shear displacements strongly exceed average compressive deformations occurring in, e.g., well-known parametric arrays. In the last decade the RF in so-called water-like media (where the shear elastic modulus is small but finite) has been intensively discussed in relation to emerging biomedical applications (e.g., Ref. 3).

At least two important problems remain unsolved. First, little is known about the effect of nonlinear wave profile distortion on the RF except in the case of a sawtooth wave with shocks.[4,5] However, in many practical cases, in particular, in biomedical research, the field can undergo significant nonlinear distortions but not reach a sawtooth stage. Second, in biological tissues and some other media, the absorption coefficient is typically linearly, rather than quadratically, dependent on frequency which cannot be described by the models with constant viscosity.

This Letter is concerned with these two problems as they apply to an arbitrary isotropic elastic medium (including a fluid). We consider the shear (solenoidal) radiation force in focused, nonlinear ultrasonic beams with different types of losses. Instead of calculating specific situations numerically we use a simplified analytical model which allows one to obtain understanding of general qualitative features of the process.

II. DISSIPATIVE RF IN A BEAM

The radiation force vector exciting period-average motions in an elastic medium is defined as

$$\mathbf{\Phi} = \left\langle \frac{\partial \sigma_{ik}}{\partial x_k} \right\rangle, \tag{1}$$

where σ_{ik} is stress tensor and angular parentheses denote period averaging. As in Ref. 1, we are most interested in a solenoidal part of the force, $\mathbf{\Phi}_s$, which induces shear displacements. Indeed, at least in fluids and water-like media shear motions evidently prevail over compressions—rarefactions. Denote the RF per unit mass as $\mathbf{F}=\mathbf{\Phi}_s/\rho$, where ρ is density. Note that in general $\mathbf{F}$ can retain a slow (as compared to the high-frequency oscillations) dependence on time and in such cases it can be considered as a dynamic force able to excite streaming in liquid and low-frequency shear wave motions in an elastic medium, including water-like materials such as biological tissues.[1]

In the case of a small-angle beam, the main component of $\mathbf{F}$ is directed along the beam axis x. The corresponding expression for F_x in terms of the components of the displacement vector $\mathbf{u}$ was obtained in Ref. 1 for a two-dimensional beam in an inhomogeneous medium. A similar expression for a three-dimensional case can be readily obtained following the same technique. For the homogeneous medium considered here, this expression reads

$$F_x \approx -Q\left[\frac{\partial}{\partial x}\left\langle\left(\frac{\partial u_x}{\partial x}\right)^2\right\rangle + \frac{\partial}{\partial x}\left\langle\frac{\partial u_x}{\partial x}\frac{\partial u_x}{\partial z}\right\rangle + \frac{\partial}{\partial y}\left\langle\frac{\partial u_x}{\partial x}\frac{\partial u_x}{\partial y}\right\rangle\right]. \tag{2}$$

Here $Q=-(\lambda+3\mu+A+2B)$, where λ and μ are Lamé coefficients and A and B are third-order Landau moduli describing quadratic nonlinearty (the third modulus, C, is not present in this approximation). In fluids and water-like media where $\mu \ll \lambda$ and $A \ll B \approx -\lambda$ (Ref. 6) this reduces to $Q \approx \lambda \approx \rho c_0^2$, where ρ_0 is density and c_0 is the longitudinal wave velocity.

a)Electronic mail: lev.a.ostrovsky@noaa.gov

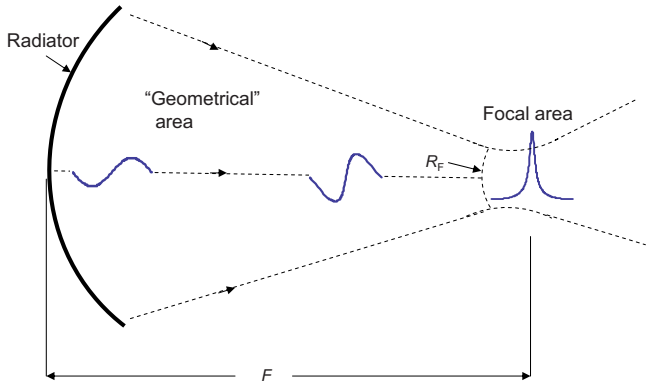


FIG. 1. (Color online) Scheme of nonlinear distortions in a focused wave. F is the focal distance of the source and R_F marks the beginning of the focal area. At $r>R_F$ and $r<R_F$, Eqs. (7) and (14) are respectively used.

Here we use the well-known Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation for a narrow-angle beam in an elastic medium and represent it in a generalized form

$$\frac{\partial M_a}{\partial X} - \frac{c_0}{2}\int\left(\frac{\partial^2 M_a}{\partial y^2} + \frac{\partial^2 M_a}{\partial z^2}\right)d\tau - \frac{\varepsilon}{c_0}M_a\frac{\partial M_a}{\partial \tau} = G(M_a, x, \tau). \tag{3}$$

Here $M_a(\tau, X) = -\partial u_x/\partial x \approx (1/c_0)\partial u_x/\partial \tau$ is the wave Mach number, G is the dissipative term, $\tau = -x/c_0$, and $X = x$ is the slowly varying coordinate, so that $|\partial/\partial X| \ll |\partial/\partial x|$. Note that, according to,[7] this equation is applicable not only to fluids but also to isotropic solids.

Multiplying Eq. (3) by M_a, taking into account that in this approximation, $cd\tau \approx dx$ and averaging over time period, after elementary integrations by parts, we have

$$\frac{1}{2}\frac{\partial}{\partial X}\left\langle\left(\frac{\partial u_x}{\partial x}\right)^2\right\rangle + \frac{1}{2}\frac{\partial}{\partial z}\left\langle\frac{\partial u_x}{\partial x}\frac{\partial u_x}{\partial z}\right\rangle + \frac{1}{2}\frac{\partial}{\partial y}\left\langle\frac{\partial u_x}{\partial x}\frac{\partial u_x}{\partial y}\right\rangle = \langle M_a G\rangle. \tag{4}$$

Comparison with Eq. (2) yields a general expression:

$$F_x(x,z) = -2Q\langle M_a G\rangle. \tag{5}$$

III. RF IN A MEDIUM WITH CLASSICAL VISCOSITY

For a fluid with a classical molecular viscosity, the dissipative term in the beam Eq. (4) has the form $G = g\partial^2 M_a/\partial\tau^2$. Here g is in general a combination of viscosity and thermal conductivity (if only the shear viscosity ν is accounted for, then $g = 2\nu/3c_0^3$). This yields

$$F_x(x,z) = 2Qg\left\langle\left(\frac{\partial M_a}{\partial \tau}\right)^2\right\rangle. \tag{6}$$

A similar expression is well known for a fluid, where $Q \approx \rho c^2$ (e.g., Ref. 3).

Consider now nonlinear evolution of a circularly symmetric focused beam (Fig. 1). To make qualitative results more transparent, we consider here the case of small dissipation when the nonlinear primary wave (without shocks) is assumed nondissipative, whereas RF is then determined taking dissipation into account. Outside the focal area the wave can be represented in the framework of a nonlinear geometrical acoustics approximation for a convergent wave:

$$M_a = M_0\frac{F}{r}\sin\left[\omega t + k(r-F) + \varepsilon M_a kr\ln\frac{F}{r}\right], \tag{7}$$

where r is the distance from the geometrical focus, F is the distance of a source from the focal point, $k = \omega/c$ is the wave number, M_0 is the wave amplitude at the source, i.e., at $r = -x = F$, and $\varepsilon = (\gamma+1)/2$ is the nonlinearity parameter (γ is the polythropic ratio).

The expression (7) is implicit but it can be integrated. Denoting $\xi = \omega t + kr - kF + \varepsilon M_a kr\ln F/r$, we have

$$\frac{\partial M_a}{\partial \tau} = M_0\frac{F}{r}\frac{\partial \xi}{\partial \tau}\cos\xi,$$

$$\frac{\partial \xi}{\partial \tau} = \omega + \varepsilon kr\ln\frac{F}{r}\frac{\partial M_a}{\partial \xi}\frac{\partial \xi}{\partial \tau} = \frac{\omega}{1 - \varepsilon kr\ln\dfrac{F}{r}\dfrac{\partial M_a}{\partial \xi}}, \tag{8}$$

so that

$$\frac{\partial M_a}{\partial \tau} = M_0\frac{\omega F}{r}\frac{\cos\xi}{1 - b(r)\cos\xi}, \tag{9}$$

where $b(r) = \varepsilon M_0 kF\ln F/r$ is the "nonlinear distance." The wave breaking and the beginning of shock formation correspond to $b = 1$, or

$$r = R_s = Fe^{-1/(\varepsilon M_0 kF)}. \tag{10}$$

Representing $d(\omega t) = d\xi(1 - b(r)\cos\xi)$, we have

$$\left\langle\left(\frac{\partial M_a}{\partial \tau}\right)^2\right\rangle = \frac{\omega^2 M_0^2 F^2}{2\pi r^2}\int_{-\pi}^{\pi}\frac{\cos^2\xi d\xi}{1 - b(r)\cos\xi} = \frac{\omega^2 M_0^2 F^2}{b^2 r^2}\left(\frac{1}{\sqrt{1-b^2}} - 1\right). \tag{11}$$

Thus, according to Eq. (6),

$$F_x = 2Qg\frac{\omega^2 M_0^2 F^2}{b^2 r^2}\left(\frac{1}{\sqrt{1-b^2}} - 1\right). \tag{12}$$

At small nonlinearity ($b^2 \ll 1$) it follows that

$$F_x = \frac{Qg\omega^2 M_0^2 F^2}{r^2} = Qg\omega^2 M_r^2, \tag{13}$$

where $M_r(r) = M_0 F/r$ is the local Mach number at radius r. For a fluid Eq. (13) is again a known expression. However, according to Eq. (12), the force rapidly increases when $b^2 \to 1$, i.e., upon steepening of the wave front. A limitation for the increase is caused by diffraction in the focal area.

IV. DIFFRACTION EFFECTS: FORCE IN THE FOCUS

In general, to estimate the diffraction stage, the full KZK Eq. (3) can be used. However, it does not allow any relevant analytical solutions, and therefore numerical calculations should be applied in each specific case. To obtain a more general result, here we use the "step-by-step" approach developed and tested in Ref. 8. It is based on breaking up the integration path into "geometrical" and "diffractive" stages

(Fig. 1). At the nonlinear stage diffraction is neglected and the quasi-spherical nonlinear solution (7) is used. At the next stage, in the focal area, diffraction plays a dominant role and nonlinearity can be neglected. These stages are matched at a radius $r=R_F$ of the order of the length of the focal area at the main frequency. For a small-angle beam $R_F \sim \lambda/\alpha^2$, where $\lambda=2\pi/k$ is the wavelength and $\alpha=a/F$ is the angular half width of the beam. Substituting it into Eq. (12) gives the maximum force, $F_x(R_F)$, at the geometrical stage if $b(R_F)<1$, i.e., the shock wave is not formed at $r>R_F$.

Let us now consider the focal area, $r<R_F$. Taking the boundary condition in the form of Eq. (7) with $r=R_F$, we obtain a linear solution in the Kirchoff approximation

$$M_a = \frac{1}{2\pi c}\int_S \frac{1}{r}\frac{\partial}{\partial \tau} M_S\left(\tau + \frac{r-F}{c}\right) ds, \tag{14}$$

where $S \approx \pi\alpha^2 R_F^2$ is the beam cross-section area at $r=R_F$ and M_S is the Mach number at $r=R_F$ defined by Eq. (7). In particular, in the focal point, $r=0$, it yields

$$M_F = \frac{\pi}{\omega}\frac{\partial M_S}{\partial \tau}\left(\tau + \frac{R_F - F}{c}\right). \tag{15}$$

From Eqs. (6) and (15) it follows for the force in the focus that

$$F_x(0) = \left(\frac{2\pi^2 Q g}{\omega^2}\right)\left\langle \left(\frac{\partial^2 M_S}{\partial \tau^2}\right)^2 \right\rangle. \tag{16}$$

Using the transforms similar to Eqs. (8) and (11), we find

$$\left\langle \left(\frac{\partial^2 M_S}{\partial \tau^2}\right)^2 \right\rangle = M_0^2 \frac{F^2\omega^4}{2\pi R_F^2}\int_{-\pi}^{\pi} \frac{\sin^2 \xi d\xi}{(1-b(R_F)\cos\xi)^5}$$

$$= M_0^2 \frac{F^2\omega^4}{2R_F^2}\frac{4+b^2(R_F)}{4(1-b^2(R_F))^{7/2}}, \tag{17}$$

so that

$$F_x = Q g M_0^2 \left(\frac{\pi\omega F}{R_F}\right)^2 \frac{4+b^2(R_F)}{4(1-b^2(R_F))^{7/2}}. \tag{18}$$

The above results are valid throughout the entire beam if $R_F > R_s$ [see Eq. (10)], i.e., if shocks are not formed. The opposite case, i.e., when a sawtooth wave is formed at $r > R_F$, has been numerically treated in, e.g., Ref. 5.

V. A "BIOLOGICAL" DAMPING LAW

Consider now another dissipative operator G that is characteristic of many solids and biological tissues at ultrasonic frequencies. Roughly, $\mathrm{Im}(k) \propto \omega$ rather than ω^2 (e.g., Ref. 9). For nonharmonic waves, this law follows from an integral representation for the dissipative term based on Hilbert's transform:

$$G = g_1 \frac{\partial H(M_a)}{\partial \tau}, \quad H(M_a) = PV\int_{-\infty}^{\infty} \frac{M_a(\tau')d\tau'}{\tau'-\tau}, \tag{19}$$

where PV denotes principal value.

In Eq. (19) H is the Hilbert integral that transposes $\sin n\omega t$ to $\cos n\omega t$ in the Fourier transform of M_a. Representing

$$M_a = \sum_{n=1}^{\infty} A_n(r)\sin(n\omega\tau) \to H(M_a)$$

$$= \pi\sum_{n=1}^{\infty} A_n(r)\,\mathrm{sign}(\omega)\cos(n\omega\tau),$$

we have

$$G = -\tau g_1 |\omega| \sum_{n=1}^{\infty} n A_n(r)\sin(n\omega\tau). \tag{20}$$

In the case of a linear harmonic wave, $M_a = a(r)\sin\omega t$, Eq. (20) yields $G=-ag_1|\omega|\sin\omega t$. Then, according to Eq. (3), for a linear plane wave $a(x)\sim\exp(-\pi g_1|\omega|x)$ as expected.

For a nonlinear periodic M_a in the form of Eq. (7), according to the known Bessel–Fubini expansion (e.g., Ref. 9), one has $A_n = (2M_0F/r)J_n(nb)/nb$ $(b<1)$. Substituting this into the expression (7) for the force, we obtain after averaging

$$F_x = (\pi/2)Qg_1(M_0F/r)^2|\omega|S_1(b);$$

$$S_1 = 4\sum_{n=1}^{\infty} n\left(\frac{J_n(nb)}{nb}\right)^2. \tag{21}$$

A simple numerical calculation shows that the sum can be well approximated by a polynomial

$$S_1 \approx 1.006 - 0.529b^2 + 3.9b^3 - 6.4b^4 + 3.625b^5. \tag{22}$$

In the linear case when $b \ll 1$, $S_1(0) \approx 1$ and in the shock formation limit, $S_1(1)=1.6$. It is remarkable that, unlike the above classical case, here nonlinear steepening of the wave does not turn the force to infinity at $b=1$ but still increases the force by a factor of 1.6.

Finally, for the field (15) in the focal point we have

$$H(M_F) = \frac{\pi R_F\alpha^2}{2c}\mathrm{sign}(\omega)\sum_{n=1}^{\infty} A_n(r) n\omega \sin(n\omega\tau),$$

$$\hat{G} = -\pi g_1 \frac{R_F\alpha^2}{c}\omega^2\left(\frac{M_0F}{R_F}\right)\mathrm{sign}(\omega)\sum_{n=1}^{\infty} n^2\left(\frac{J_n(nb)}{nb}\right)^2, \tag{23}$$

so that

$$F_x = Qg_1\pi|\omega|^3\left(\frac{M_0F}{R_F}\right)^2\left(\frac{R_F\alpha^2}{2c}\right)^2 S_2(b),$$

$$S_2 = 4\sum_{n=1}^{\infty} n^3\left(\frac{J_n(nb)}{nb}\right)^2 \approx 4\left(\frac{0.415-0.567b+0.231b^2}{1.819-3.673b+1.856b^2}\right). \tag{24}$$

The latter approximation of S_2 is valid, with an error of a few percent, for up to $b \leqslant 0.97$.

VI. ESTIMATES

For estimates we take $F=10$ cm, $a=3$ cm, $\varepsilon=3.5$, and $M_0=5\cdot10^{-5}$ (in water the latter corresponds to pressure amplitude at the transducer of approximately 100 kPa). These

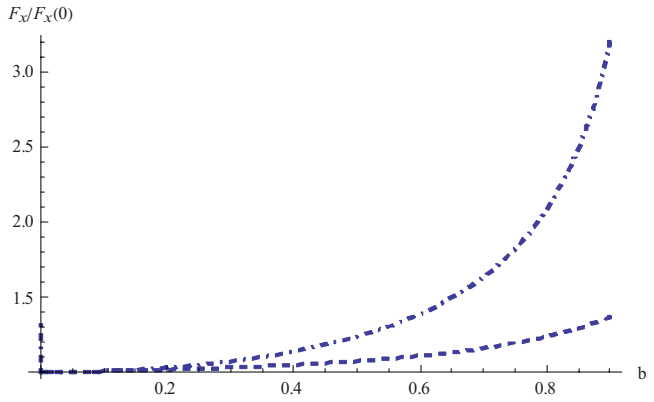


FIG. 2. (Color online) Increase of the dissipative RF (normalized to its quantity at $b=0$) at the beam axis with the nonlinearity parameter b. Dash-dotted line corresponds to classical viscosity, dashed line to "biological" damping (damping rate proportional to frequency).

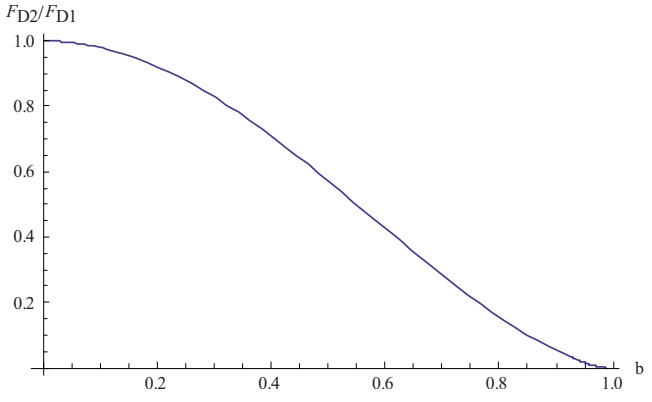


FIG. 3. (Color online) Ratio of the radiation force, F_{D2}, for "biological" damping, (24) to the force, F_{D1}, for the classical damping, (18), in the focal point.

quantities are among those that are typical in biomedical applications. In this case, at frequencies below 4.2 MHz, $R_s < R_F$, so that shocks do not form up to the diffraction zone, and the above theory is valid in the entire area.

Next the effect of nonlinear distortions on the dissipative force is evaluated. Figure 2 shows the increase of the "geometrical" forces (12) and (21) with the nonlinearity parameter b. In a classical viscous medium this increase is, in a geometrical approximation, unlimited, whereas in the "biological" case it is limited.

Figure 3 depicts the ratio of the force values in the focus in these two cases at equal attenuation coefficients for linear wave at the main frequency ω. Here again, the force increases much stronger in a "classical" viscous medium.

VII. CONCLUSIONS

Radiation force, which is responsible for the generation of streaming in fluids and shear displacements in elastic media including water-like ones (such as biological tissues), is crucially dependent on the rate of nonlinear distortions in the primary ultrasonic beam, as well as on the dissipation mechanisms in the medium. In water and other media with classical molecular viscosity, nonlinearity in the primary beam dramatically (at the geometrical stage infinitely) increases radiation force. For losses proportional to frequency, which is characteristic of many biological objects, this increase is not as strong but still significant.

The above nonlinear model which neglects attenuation in the primary beam is certainly not intended to replace detailed numerical computations for specific cases. On the other hand, it seems to be universal enough to obtain qualitative understanding of the main physical processes and in some cases provide a realistic analytical (or at least semianalytical) quantitative description of experimental situations.

This approach can be extended to still more intensive beams when shock waves are formed before the focal area. It can be shown that in a fluid with classical viscosity defining the width of a shock front, radiation force is proportional to M_a^3 which agrees with the result obtained in Ref. 4; in the "biological" case the corresponding results can be obtained using the results of Ref. 10, where the propagation of shock waves in media with power law of dissipation is considered. Note also that a similar radiation force, at least for classical viscous losses, drives the thermal action of the sound,[11] and the above analysis may be of use in that case as well.

After the radiation force is found, calculation of shear displacements (or streaming in the case of a fluid) driven by the RF can be readily performed from the averaged dynamic equations similarly to that done in Ref. 1 for a quasi-harmonic beam.

ACKNOWLEDGMENT

The author is grateful to O. A. Sapozhnikov and A. P. Sarvazyan for valuable discussions.

[1] L. A. Ostrovsky, A. M. Sutin, Yu. A. Il'inskii, O. V. Rudenko, and A. P. Sarvazyan, "Radiation force and shear motions in inhomogeneous media," J. Acoust. Soc. Am. **121**, 1323–1331 (2007).
[2] C. Eckart, "Vortices and streams caused by sound waves," Phys. Rev. **73**, 68–86 (1948).
[3] O. V. Rudenko and A. P. Sarvazyan, "Nonlinear acoustics and biomedical applications," Crit. Rev. Biomed. Eng. **3**, 6–19 (2000).
[4] Yu. A. Pishchalnikov, O. A. Sapozhnikov, and T. V. Sinilo, "Increase in the efficiency of the shear wave generation in gelatin due to the nonlinear absorption of a focused ultrasonic beam," Acoust. Phys. **48**, 214–219 (2002).
[5] S. J. Younghouse, "Acoustic streaming at high Reynolds numbers produced by focused sound beams with shocks in real fluids," Master's thesis, University of Texas at Austin, Austin, TX, 1998.
[6] M. F. Hamilton, Yu. A. Il'inskii, and E. A. Zabolotskaya, "Separation of compressibility and shear deformation in the elastic energy density," J. Acoust. Soc. Am. **116**, 41–44 (2004).
[7] E. A. Zabolotskaya, "Sound beams in a nonlinear isotropic solid," Sov. Phys. Acoust. **32**, 296–299 (1986).
[8] L. A. Ostrovsky and A. M. Sutin, "Focusing of finite-amplitude acoustic waves," Sov. Phys. Dokl. **20**, 275–277 (1975).
[9] H. A. H. Jongen, J. M. Thijssen, M. van den Aarssen, and W. A. Verhoef, "A general model for the absorption of ultrasound by biological tissues and experimental verification," J. Acoust. Soc. Am. **79**, 535–540 (1986).
[10] S. S. Kashsheeva, O. A. Sapozhnikov, V. A. Khokhlova, M. A. Averkiou, and L. A. Crum, "Nonlinear distortion and attenuation of intense acoustic waves in lossy media obeying a frequency power law," Acoust. Phys. **46**, 170–177 (2000).
[11] A. A. Karabutov, O. V. Rudenko, and O. A. Sapozhnikov, "Theory of thermal self-focusing with allowance for the generation of shock waves and acoustic streaming," Sov. Phys. Acoust. **34**, 371–374 (1988).

Comparison of the Faran Cylinder Model and the Weak Scattering Model for predicting the frequency dependence of backscatter from human cancellous femur *in vitro* (L)

Keith A. Wear[a)]
Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland 20993

Frederic Padilla and Pascal Laugier
UPMC University of Paris 06, UMR 7623, LIP, F-75005, Paris, France and CNRS, UMR 7623, Laboratoire d'Imagerie Paramétrique, F-75006, Paris, France

(Received 2 May 2008; revised 11 June 2008; accepted 12 June 2008)

This letter presents the first side-by-side comparison of the Faran Cylinder Model and the Weak Scattering Model for predicting backscatter from human femur. Both models are applied to the same dataset of frequency-dependent backscatter coefficients from 26 human femur cancellous bone samples *in vitro*. The Faran Cylinder Model predicts a slightly slower rate of increase of backscatter with frequency than the Weak Scattering Model, but both models are in reasonable agreement with the data and with each other, given the uncertainty in the measurements.
 [DOI: 10.1121/1.2956480]

PACS number(s): 43.80.Qf, 43.80.Jz, 43.80.Ev [FD] Pages: 1408–1410

I. INTRODUCTION

Two models for backscatter from cancellous bone have been extensively investigated: the Faran Cylinder Model (Wear, 1999) and the Weak Scattering Model (Chaffai *et al.*, 2000, Jenson *et al.*, 2003). The Faran Cylinder Model portrays cancellous bone as a network of cylinders (trabeculae). Faran's theory of scattering from cylinders is used to predict scattering from the trabecular network (Faran, 1951). The Weak Scattering Model is based on a random description of the scattering tissue. It uses the structural autocorrelation function of a cancellous bone sample (e.g., measured from a microCT scan or modeled by an analytical function) as an input to predict backscatter.

There are advantages and disadvantages to the two models. Unlike the Weak Scattering Model, the Faran Cylinder Model does not require isotropy or low contrast in acoustic properties between trabeculae and the fluid filler (water or marrow). However, the Weak Scattering model has far more flexibility than the Faran Cylinder Model to accommodate complex cancellous bone microstructure. Both models assume incoherent backscatter and negligible multiple scattering. For a more detailed comparison of these models, see Wear (2008).

Despite the considerable differences in the two models, both have been shown to accurately predict the dependencies of backscatter on frequency (Wear, 1999; Chaffai *et al.*, 2000; Jenson *et al.*, 2003; Padilla *et al.*, 2003) and trabecular thickness (Wear and Laib, 2003; Jenson *et al.*, 2003; Padilla *et al.*, 2003; Padilla *et al.*, 2006) in human calcaneus *in vitro*. In addition, the Weak Scattering Model has been shown to accurately predict the dependencies of backscatter on frequency (Padilla *et al.*, 2006) and trabecular thickness (Padilla *et al.*, 2006) in human femur *in vitro*.

This letter reports the first side-by-side comparison of predictions based on the Faran Cylinder and Weak Scattering Models with data from human cancellous femur samples *in vitro*.

II. METHODS

A. Bone specimens

Twenty six cancellous bone specimens were obtained from fresh cadavers as described previously (Jenson *et al.*, 2006). Ethics approval for collection of samples was granted by the Human Ethics Committee of the Institute of Anatomy at the University René Descartes (Paris, France). The donors or their legal guardians provided informed written consent to provide their tissues for investigation in accordance with legal clauses stated in the French Code of Public Health (Code de la Santé Publique Français). Briefly, each femur was excised and the soft tissues removed manually. Then the specimens were immediately stored at −20 °C. The femoral head and the diaphysis were first removed, leaving the proximal part of the femur (neck and trochanter). Slices of trabecular bone with parallel faces and thickness ranging from 0.85 to 1.35 cm were cut from the proximal femur in the plane defined by the cephalo-caudal and medio-lateral directions using a circular saw under continuous irrigation, leading to an orientation of the trabeculae mainly in planes parallel to the faces. Each sample's thickness was measured using an electronic caliper. The specimens were defatted using a dichloromethane ($C_2H_2Cl_2$) solution (immersion for two or three days) and water jet. To remove any air bubbles

[a)]Author to whom correspondence should be addressed. Electronic mail: kaw@fda.hhs.gov

trapped in the defatted samples, they were vacuum degassed under water in a desiccator before ultrasonic measurements.

B. Ultrasonic measurements

Ultrasonic measurements were performed as described previously (Jenson *et al.*, 2006). Briefly, bone samples were immersed in water, and through-transmission measurements were performed using a pair of focused broadband transducers (1 MHz center frequency, 29 mm diameter, 35 mm focal length, V391, Panametrics Inc., Waltham, MA). Emission and reception of the signals were performed using a plug-in ultrasonic pulser receiver, amplifier, and digitizer (SFT 4001H PCI, Sofratest, Ecquevilly, France). The transducers were mounted coaxially, separated by twice the focal length. The specimen was placed in the focal zone. Transmitted and backscattered radio-frequency signals were recorded along two-dimensional (2D) scans in steps of 1 mm (Motion Controller, MM4006, Newport, Irvine, CA). Radio frequency signals were amplified, time averaged, and digitized at 60 MHz. The ratio of the Fourier transform of the signal transmitted through the sample to that of a reference signal transmitted through water was computed. The log magnitude of the ratio divided by the sample thickness gave the frequency-dependent attenuation coefficient. The slope of a linear regression fit to the attenuation coefficient between 0.4 and 0.8 MHz gave the attenuation slope.

Backscatter measurements were performed using a substitution method as described previously (Chaffaï *et al.*, 2000). Briefly, a reference echo was acquired on a plane reflector (steel plate) placed at a distance equal to the position of the scattering volume. Then, an echo signal was acquired from the scattering of the incident pulse onto a bone specimen. This signal was time weighted using a Hamming function in order to keep only the part of the signal backscattered from a volume approximately 8 mm in length placed in the center of the specimen. Then the backscatter coefficient was calculated by computing the ratio of the frequency power spectrum of the time-gated echo signal to the power spectrum of the reference signal. Corrections were made to compensate for attenuation, Hamming gate function, and diffraction (Chaffaï *et al.*, 2000).

C. Synchrotron microtomography

Cancellous bone samples were investigated using synchrotron radiation at the European Synchrotron Radiation Facility, France, as described previously (Bossy *et al.*, 2005). Briefly, three-dimensional (3D) reconstructions of the internal structures were obtained from sets of 2D projections under different angles of view. The voxel size was 10 μm $\times$10 μm$\times$10 μm. The acquisition setup was based on 3D parallel tomography and has been previously described in detail (Salome *et al.*, 1999). Three-dimensional reconstructions were processed in order to extract morphological parameters such as mean trabecular thickness (Hildebrand and Rüegsegger, 1997).

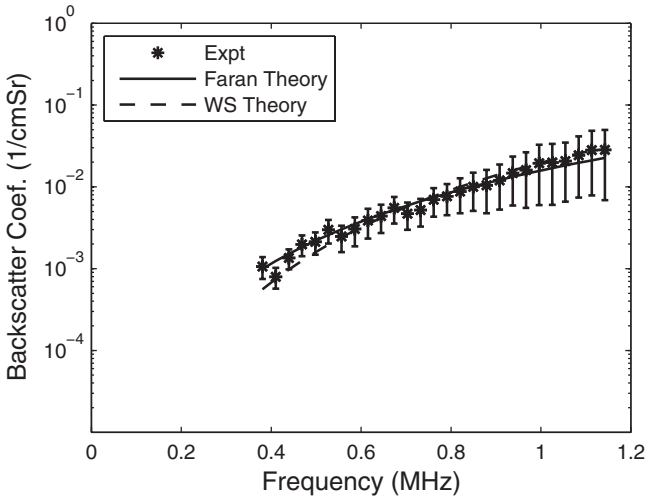


FIG. 1. Average backscatter coefficient vs. frequency—experimental and theoretical values.

D. Data analysis

The experimental backscatter coefficient data, $\eta(f)$, for each cancellous bone specimen was fit to a power law of frequency, $\eta(f)=Af^n$ over the range from 0.4 to 1.2 MHz. The frequency dependence of backscatter was characterized by the exponent n. For an infinite, thin (compared to the wavelength) cylinder, $n=3$. For a point scatterer, $n=4$. The Faran Cylinder Model prediction (as a function of frequency) for each cancellous bone specimen was generated based on the trabecular thickness value obtained from microCT as described previously (Wear, 1999; Wear and Laib, 2003). The Weak Scattering Model prediction (as a function of frequency) for each cancellous bone specimen was generated based on the microCT dataset as described previously (Jenson *et al.*, 2003: Padilla *et al.*, 2003; Padilla *et al.*, 2006).

III. RESULTS

Figure 1 shows experimental backscatter coefficient vs. frequency. The average model fits to the data are also shown. Both models fit the data well. The Faran Cylinder Model has a slower rate of increase with frequency and may fit the data better at lower frequencies. The Weak Scattering Model has a faster rate of increase with frequency and may fit the data better at higher frequencies. Table I shows the power law fit parameters.

The average Faran Cylinder Model exponent (n), 2.84, was slightly less than 3 as expected for a thin cylinder with a diameter much smaller than a wavelength (Wear, 1999; Chaffai *et al.*, 2000). The mean trabecular thickness for the

TABLE I. Backscatter coefficient power law fits, $\eta(f)=Af^n$, for experimental data, Faran Cylinder Model, and Weak Scattering Model. Frequency f is measured in Hz, and backscatter coefficient $\eta(f)$ is measured in $cm^{-1}\,Sr^{-1}$.

	A	Exponent (n)	95% Confidence Interval for n
Experimental	2.92×10^{-21}	3.13	2.96–3.30
Faran cylinder	1.41×10^{-19}	2.84	2.82–2.86
Weak scattering	3.25×10^{-24}	3.63	3.59–3.67

samples used in this study was 132 μm microns, which is much smaller than the wavelength at 1 MHz in water (approximately 1.5 mm).

IV. CONCLUSION

The letter presents the first side-by-side comparison of the Faran Cylinder Model and the Weak Scattering model for predicting the frequency dependence of backscatter from human cancellous femur *in vitro*. Both models fit the data well. Measurements of backscatter coefficient in cancellous bone are extremely difficult, and their accuracy and precision are limited by many factors including small sample size, sample inhomogeneity, phase cancellation, coherent effects, limited bandwidth, and multiple scattering (Wear, 2001). Since the accuracy of measurements of the frequency dependence of scattering (i.e., exponent of a power law fit) on cancellous bone samples in experiments such as reported here cannot be expected to be better than a few tenths, it is not possible, based on this data, to conclude that either model performs better than the other. Moreover, both models are sufficiently simplistic in their assumptions that neither can be expected to model the physics extremely accurately. The main conclusion to draw from this side-by-side comparison is that the true exponent for cancellous femur is probably near the 3.0–3.5 range and that both models are in reasonable agreement with this. The Faran Model tends to predict a slower rate of increase of backscatter with frequency and may fit the data better at low frequencies while the Weak Scattering Model tends to predict a faster rate of increase of backscatter with frequency and may fit the data better at high frequencies. One possible future application of these models is their extension to predict the total scattering (that is, scattering integrated over all angles) component of broadband ultrasound attenuation, which (along with speed of sound) is one of the two primary clinical quantitative ultrasound measurements for bone.

Bossy, E., Padilla, F., Peyrin, F., and Laugier, P. (**2005**). "Three-dimensional simulation of ultrasound propagation through trabecular bone structures measured by synchrotron microtomography," Phys. Med. Biol. **50**, 5545–5556.

Chaffai, S., Roberjot, V., Peyrin, F., Berger, G., and Laugier, P. (**2000**). "Frequency dependence of ultrasonic backscattering in cancellous bone: Autocorrelation model and experimental results." J. Acoust. Soc. Am. **108**, 2403–2411.

Faran, J. J. (**1951**). "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–411.

Hildebrand, T., and Rüegsegger, P. (**1997**). "A new method for the model independent assessment of thickness in three-dimensional images," J. Microsc. **187**, 67–75.

Jenson, F., Padilla, F., and Laugier, P. (**2003**). "Prediction of frequency-dependent ultrasonic backscatter in cancellous bone using statistical weak scattering model," Ultrasound Med. Biol. **29**, 455–464.

Jenson, F., Padilla, F., Bousson, V., Bergot, C., Laredo, J. D., and Laugier, P. (**2006**). "*In vitro* ultrasonic characterization of human cancellous femoral bone using transmission and backscatter measurements: Relationships to bone mineral density," J. Acoust. Soc. Am. **119**, 654–663.

Padilla, F., Peyrin, F., and Laugier, P. (**2003**). "Prediction of backscatter coefficient in trabecular bones using a numerical model of three-dimensional microstructure," J. Acoust. Soc. Am. **113**, 1122–1129.

Padilla, F., Jenson, F., and Laugier, P. (**2006**). "Estimation of trabecular thickness using ultrasonic backscatter," Ultrason. Imaging **28**, 3–22.

Salome, M., Peyrin, F., Cloetens, P., Odet, C., Laval-Jeantet, A. M., Baruchel, J., and Spanne, P. (**1999**). "A synchrotron radiation micotomography system for the analysis of trabecular bone samples," Med. Phys. **26**, 2194–204.

Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: Theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.

Wear, K. A. (**2001**). "Fundamental precision limitations for measurements of frequency dependence of backscatter: Applications in tissue-mimicking phantoms and trabecular bone," J. Acoust. Soc. Am. **110**(6), 3275–3282.

Wear, K. A., and Laib, A. (**2003**). "The relationship between ultrasonic scattering and microarchitecture in human calcaneus," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 979–986.

Wear, K. A. (**2008**). "Ultrasonic scattering from cancellous bone: A Review," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1432–1441 (2008).

Educational ultrasound nondestructive testing laboratory

Vladimir Genis[a] and Michael Zagorski[b]
Drexel University, One Drexel Plaza, Philadelphia, Pennsylvania 19104

(Received 29 February 2008; accepted 10 June 2008)

The ultrasound nondestructive evaluation (NDE) of materials course was developed for applied engineering technology students at Drexel University's Goodwin College of Professional Studies. This three-credit, hands-on laboratory course consists of two parts: the first part with an emphasis on the foundations of NDE, and the second part during which ultrasound NDE techniques are utilized in the evaluation of parts and materials. NDE applications are presented and applied through real-life problems, including calibration and use of the latest ultrasonic testing instrumentation. The students learn engineering and physical principles of measurements of sound velocity in different materials, attenuation coefficients, material thickness, and location and dimensions of discontinuities in various materials, such as holes, cracks, and flaws. The work in the laboratory enhances the fundamentals taught during classroom sessions. This course will ultimately result in improvements in the educational process ["The greater expectations," national panel report, http://www.greaterexpectations.org (last viewed February, 2008); R. M. Felder and R. Brent "The intellectual development of Science and Engineering Students. Part 2: Teaching to promote growth," J. Eng. Educ. **93**, 279–291 (2004)] since industry is becoming increasingly reliant on the effective application of NDE technology and the demand on NDE specialists is increasing. NDE curriculum was designed to fulfill levels I and II NDE in theory and training requirements, according to American Society for Nondestructive Testing, OH, Recommended Practice No. SNT-TC-1A (2006). © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2966175]

PACS number(s): 43.10.Sv, 43.20.Ye, 43.35.Yb, 43.40.Le [TDR] Pages: 1411–1418

I. INTRODUCTION

For more than four years, Drexel University's Goodwin College has been offering the Applied Engineering Technology (AET) major to full- and part-time students at Drexel University and also as a dual enrollment program to the students of Delaware County Community College and Burlington County College. The AET program is based on a cyclic model of the relationship between knowledge production and improvement of practice in undergraduate education and clearly distinguishes itself from traditional engineering programs in the following ways.

- It creates a bridge between the engineer/scientist and the technical and/or production workforce.
- The curriculum places emphasis on the application of theory rather than on derivations and proofs.
- The majority of courses are fully integrated with training and laboratory experience, including extensive use of software and industrial case studies.
- Faculty members with extensive industrial and academic experience support the program.

To achieve the AET program's goals and improve students' learning, Goodwin College is in the process of expanding and upgrading its educational facilities. In order to facilitate this process, several state-of-the-art laboratories were created to provide hands-on education and training to students who are pursuing a degree in AET with a concentration in electrical, mechanical, or manufacturing engineering technology. AET laboratories are used extensively for most of the program's courses, including the newly developed course in Ultrasound Nondestructive Evaluation (NDE) of Materials. This course consists of two parts: the first part with an emphasis on the foundations of NDE, and the second part, during which NDE techniques of parts and materials are presented and applied through real-life problems. Specifically, the students learn the engineering and physical principles of measurements of sound velocity in different materials, attenuation coefficients, directivity pattern of ultrasonic transducers, and location and dimensions of various discontinuities, such as holes, cracks, cavities, and flaws. The work in the laboratory enhances the fundamentals taught in the classroom sessions. The purpose of this project is not to develop NDE techniques, but to expose students to existing techniques by introducing them to the equipment and methods used in real-world industrial applications. The outcomes of this project will lead to improvements in the education process, since industry is becoming increasingly reliant on the effective application of NDE technology, and the demand on the NDE technicians is increasing.

NDE is one of the most powerful and cost-effective techniques for quality and safety control of structures, parts, and products. NDE of materials and components is crucial to aerospace, naval, railroad, and other industries. For example, aerospace designers and manufacturers are frequently faced with the need to validate the integrity of structural parts for military and commercial aircrafts.[1] Transportation equipment is highly specialized and safety sensitive; therefore, it is important that all aspects of a failure are investigated, not only

[a]Electronic mail: genisv@drexel.edu
[b]Electronic mail: mz24@drexel.edu

for the purpose of classifying the failure mode but also to determine its cause from an engineering mechanics and design point of view. Similar issues exist in other areas, such as maintenance and diagnostic techniques for nuclear power plants and petrochemical industry. The NDE course was designed in response to the requirements of local and regional industries in need of such techniques, and more importantly, of qualified personnel. Due to its distinctive nature, Drexel's AET program is uniquely positioned to incorporate NDE courses and experiences as part of its curriculum, thus creating the necessary knowledge and experience among its students and graduates.

II. COURSE AND LABORATORY DEVELOPMENT

The developed ultrasound NDE educational laboratory will serve as a training center for undergraduate AET students as well as for the workforce of companies, such as Boeing, Lockheed Martin, and PECO Energy, with whom Drexel has a rich history of educational and research partnership. After careful consideration and discussions with the largest employers in the Atlantic region and representatives of the American Society for Nondestructive Testing, along with our own research, educational, and engineering experience, we came to the conclusion that the creation of a unique ultrasound NDE laboratory would significantly benefit our students and working engineering personnel. Such educational laboratories are almost nonexistent in the tristate (PA, NJ, and DE) geographic area and would be welcomed by the working community. The establishment of a state-of-the art laboratory for NDE purposes will allow Drexel and its community-college partners to develop training options for technicians located in the region's key industries. NDE curriculum is designed to fulfill levels I and II NDE in theory and training requirements, according to ASNT Recommended Practice No. SNT-TC-1A (2006).

During the development of this course, representatives from several companies enthusiastically expressed their interest in participation in this project. Specifically, Mr. Rich Bottari, manager of Stork Materials Technology, and Mr. Tony Midora of Precision Measurement Technology agreed to provide the guidance for designing the experiments and conducting evaluations of the students' competence. The simulation of the ultrasound NDE applications used by companies in industry was implemented in the project. Moreover, the industry partners suggested providing the Certificate of Qualification to the students who demonstrated the level of competence required by industry (ASNT Recommended Practice No. SNT-TC-1A, 2006). Also, the AET program's Industrial Advisory Committee, which is represented by Mr. Mark Carlos, the president of Physical Acoustics, Inc., made necessary recommendations regarding the course procedures, reports, and evaluation. Based on these recommendations, evaluation reports, and student course evaluation forms, the necessary changes in the course guidance and laboratory procedures will be implemented. In addition, the industrial partners, in collaboration with the Drexel's faculty, developed real-world industrial problems and provided instruction during the laboratory sessions (Fig. 1).

FIG. 1. (Color online) Guided instruction is provided during the laboratory sessions by qualified industry collaborators.

The suggested laboratory experiments were presented by the industrial partners to the faculty responsible for the final phrasing of the given task, such that the description of the task, including the objective of the project, its motivation, and the expected deliverables with the timeline were clear and understood by the students. The industrial collaborators also agreed to provide samples of parts and materials used for inspection (Fig. 2). For example, PMT, Inc., the local representative of GE Inspection Technology, donated three parts of large diameter bearings, which are used for calibration of flaw detectors in the NDE procedures. Sonaspection International, Inc. supplied the welding samples, which are used for detection and characterization of the weld's defects. Westech Inspection, Inc. provided the sample of a portion of a pipe covered by corrosion.

Desired learning outcomes for students who complete this course are as follows.

- Students will become familiar with basic ultrasonic instrumentation, gain hands-on experience with ultrasonic and electronic equipment, and be able to demonstrate the basic principles of ultrasound imaging.
- Students will be able to use an ultrasonic inspection system for quality control analysis of various materials incorporating the latest development in imaging software technology, ultrasonic instrumentation, and mechanical positioning hardware assemblies.
- Students will become familiar with the practical methods and techniques of NDE used in industry and improve their problem-solving capabilities.
- Students will carry out laboratory procedures and techniques with guidance from an instructor and describe the purpose of the procedures, techniques, and the results in their laboratory reports.

During the laboratory sessions, the students learn the engineering and physical principles of measurements of

FIG. 2. (Color online) Large diameter bearing (left) and welding samples (right).

sound velocity in different materials, attenuation coefficients, directivity pattern of ultrasonic transducers, and location and dimensions of heterogeneities in various materials, such as flaws, cavities, layers, and holes. The industrial case studies in laboratory environment enhance the fundamentals taught in the classroom sessions. This approach will provide an excellent introduction to the manufacturing environment for the undergraduates, develop project leadership skills, and facilitate the development of teamwork.

III. LABORATORY PROCEDURES

The three-credit, quarter-based (11 weeks) NDE course consists of lecture and laboratory work each week. Laboratories are organized around current developments in the field of ultrasound NDE of materials. During the laboratory sessions, students are introduced to tools, methodologies, and techniques that may be useful to solving the problem. Finally, students carry out experiments, evaluate their results using various methods and techniques, and describe the results of the experiments in individual reports for each laboratory session. After completion of all laboratory sessions, each team is responsible for writing a final report that summarizes the current state in the area, describes the experimental techniques utilized, discusses the expected outcomes, provides data of the actual outcomes, and explains the reasons for the departures between the expected and the actual results. The teams analyze the data, draw conclusions, and suggest possible ways for improving the accuracy of their experiments. The teams then present their findings to the class as a whole. The following experiments are carried out during the laboratory sessions.

A. Measurements of the sound velocity in water

Most applications of NDE of materials rely on accurate measurements of the sound velocity in water and other materials. The basic principle of sound velocity determination is to measure the time between transmitted and received ultrasound signals (the time of flight).[2] In these experiments, the measurement of the transducer displacement is more convenient and accurate than the measurement of the transmitter/receiver or transducer/reflector distance. Such a technique allows one to eliminate additional artifacts caused by the time delays from the transducers and attached electronics. During this laboratory session, the two sets of measurements are carried out: (1) Using one transducer, which is used as a transmitter/receiver, and the reflector, and (2) using two transducers, where one of them is used as a transmitter and the other one is used as a receiver. The schematic of the experimental setup is shown in Fig. 3. The placement of the transducers is controlled by the LABVIEW virtual instrument (VI).

All measurements are conducted in the far field. Initially, transducers are perfectly aligned in X and Y directions. Then, one of the transducers is moved in Z direction by 50 mm increments to three different positions and the data are recorded. By measuring the time interval between two consecutive pulses and knowing the displacement of one of the transducers, one can estimate the sound velocity in water,

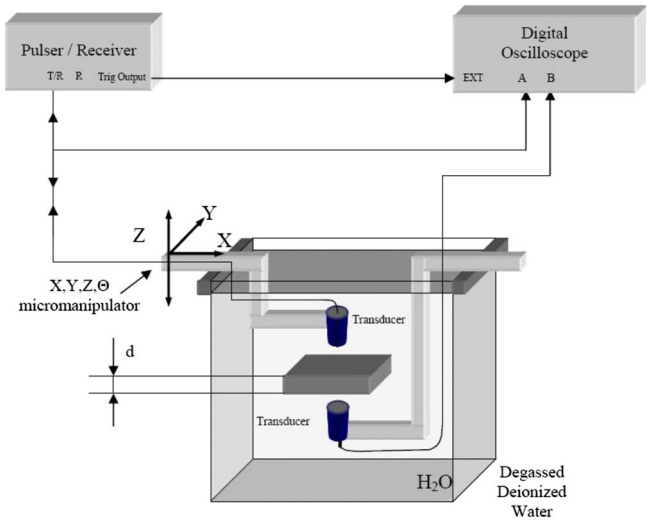


FIG. 3. (Color online) Measurements of the sound velocity in water and other materials.

as $c=z/\Delta t$, where z is the displacement of the transducer and Δt is the time interval between two consecutive received pulses. If one of the transducers is replaced with the reflector, then the sound velocity can be calculated as $c=2z/\Delta t$, since the ultrasound wave propagates in both directions—from the transducer, which acts as a transmitter, to the reflector and then back to the transducer, which acts as a receiver. Collected data are then transferred to the computer under control of LABVIEW VI and saved as a text file for future processing using Microsoft Excel. The results obtained using both techniques are later averaged.

The example of the stored and retrieved data of the received ultrasonic signal is presented in Fig. 4.

B. Measurements of the sound velocity in other materials

The propagation velocity of traveling waves is a characteristic of the media in which they travel and is generally not dependent on the other wave characteristics such as frequency, period, and amplitude.[3] The measurement system is immersed in a water tank and consists of a pair of piezoelectric transducers of similar frequency characteristics. The

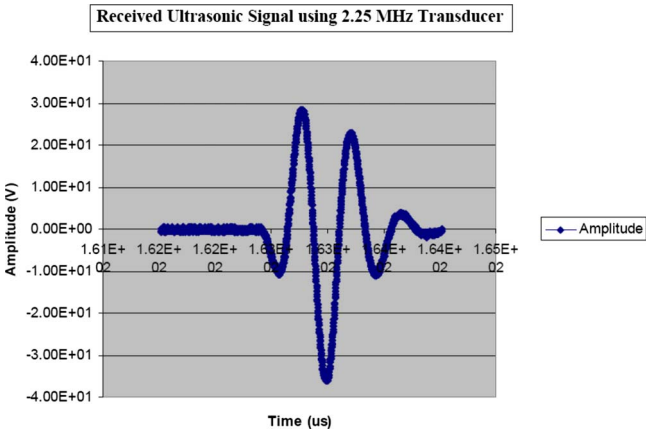


FIG. 4. (Color online) Received ultrasound signal.

Plexiglas plate and two various rubber plates are used for determining the sound velocity in these materials using the through-transmission configuration (Fig. 3). The distance between transducers is larger than the thickness of the sample material, allowing a free alignment and positioning of the sample along the ultrasonic beam. Two methods for measurements of the sound velocity in the samples are used.

Method 1. The arrival time of the received signal, τ_1, is recorded without placing the sample material between the transducers. Then, the sample material is placed between the transducers and the arrival time of the received signal, τ_2, is recorded again. The sound velocity in the sample material can now be easily determined via the following expression:

$$c_m = \frac{d}{\frac{d}{c} - \Delta\tau}, \tag{1}$$

where c is the sound velocity in water, which was previously determined, d is the thickness of the sample material, and $\Delta\tau = \tau_1 - \tau_2$.

Method 2. The arrival time of the received signal, τ_1, is recorded without placing the sample material between the transducers. Then, the sample material is placed between the transducers and the receiver is moved back from the transmitter by the distance, which is equal to the thickness of the sample material. The arrival time of the received signal, τ_2, is recorded again, and the sound velocity is determined by the following expression:

$$c_m = \frac{d}{\Delta\tau}, \tag{2}$$

where $\Delta\tau = \tau_2 - \tau_1$.

It is worth mentioning that method 2 does not require the knowledge of the sound velocity in water, since the water path for both measurements remains the same and is eliminated in the calculations. In the final report, students should clearly state the advantages and disadvantages of both methods and compare the results obtained by both methods. The described above technique can be also used for measuring the thickness of the sample materials as well as the dimensions of discontinuities in the sample materials.

C. Directivity pattern measurements

The directivity pattern is an important far-field characteristic of an ultrasonic transducer.[4,5] The directivity pattern is a dimensionless and relative parameter. The mathematical expression for the normalized directivity pattern of the plane circular piston is

$$p(\theta) = \frac{2J_1(ka \sin\theta)}{ka \sin\theta}, \tag{3}$$

where $k = 2\pi f/c$ is the wave number, J_1 is the first order Bessel function, a is the radius of the transducer, and c is the sound velocity in water. The geometry used in deriving the far-field radiation characteristics of a circular plane piston is presented in Fig. 5.

In this laboratory session, the directivity pattern is determined analytically and also measured experimentally. During the experiment, a projector and a hydrophone are separated by the minimum acceptable distance, x, to minimize interference from reflections. The standard criteria for uniform circular pistons are

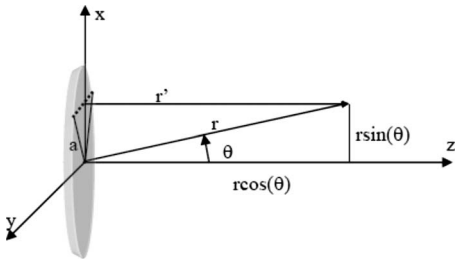


FIG. 5. Far-field pressure distribution of a circular plane piston.

$$x \geq \frac{d^2}{4\lambda}, \tag{4}$$

where $\lambda = c/f$ is the wavelength and f is the resonant frequency of the transducer.

The experimental setup for measuring the directivity pattern is similar to that of the one presented in Fig. 3, with the sample material between the transducers removed. Initially, both transducers are aligned in x, y, and θ directions. The alignment is based on the recording of the maximum amplitude of the received signal. Directivity pattern measurements are carried out with the hydrophone (receiver) fixed and the transmitter rotated using LABVIEW VI from 0° to 10° in 0.1° increments. For each angle, the peak-to-peak voltage of the received signal is recorded, and the data are saved in the computer as text files. Three pairs of transducers, 12.7 mm in diameter, having resonant frequencies of 2.25, 3.5, and 5 MHz are used in these experiments. The results of the computed and experimentally obtained directivity pattern for the 3.5 MHz resonant frequency transducer are presented in Fig. 6.

D. Measurements of the attenuation coefficient of the ultrasonic waves

The attenuation of a wave is frequency dependent and is determined by scattering and absorption, which are properties of the medium through which the wave passes.[6,7] In this laboratory session, the immersion technique is used for the measurements of the attenuation coefficient in Plexiglas. Three Plexiglas plates, which have the same mechanical and

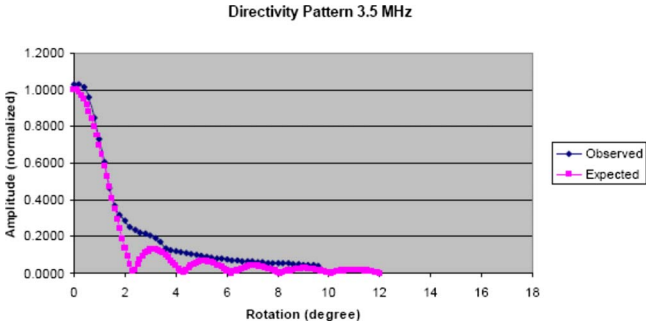


FIG. 6. (Color online) Computed and measured directivity patterns.

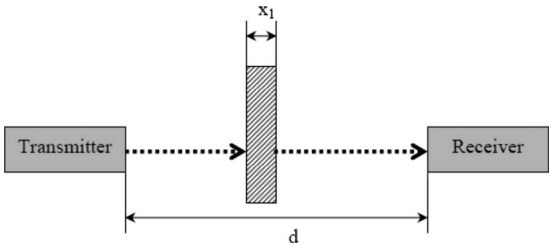


FIG. 7. Schematic of the attenuation coefficient measurement (x_1).

physical properties and thicknesses of $x_1=0.58$ cm, $x_2=1.19$ cm, and $x_3=1.69$ cm, respectively, are used in the experiments. The experimental setup is presented in Fig. 3. As can be seen, the Plexiglas pate is placed in the path of a propagating ultrasonic wave, which is partially reflected by changes in the acoustic impedance along its path. At least two samples of the Plexiglas are required for the experiment, since the reflection coefficient of the Plexiglas is not known and should be eliminated from the evaluation of the attenuation coefficient. Both transducers having the same resonant frequency are preliminary aligned in x, y, and θ directions. The alignment is verified by the maximum peak-to-peak received signal.

The following procedure is used for the measurements of the attenuation coefficient.

(1) The Plexiglas plate of the thickness x_1 is placed between the transducers (Fig. 7).
The peak-to-peak voltage of the received signal is recorded by the digital oscilloscope and saved in the computer.
(2) The Plexiglas plate x_1 is replaced with the Plexiglas plate x_2 (Fig. 8).

To keep a constant length of the water path and eliminate the changes of attenuation in the water, the distance between the transducers is increased by x_2-x_1. This change could be done by moving either the transmitter or the receiver away from each other. The time wave form of the received signal is then stored in the computer.

The attenuation coefficient is computed as follows:

$$\alpha_f = \frac{20}{x_2 - x_1}(\log_{10} V_{\text{p.p.-}x_1} - \log_{10} V_{\text{p.p.-}x_2}) \text{ dB/cm}. \quad (5)$$

The procedure is repeated when the Plexiglas plate x_2 is replaced with the plate x_3. Then, the distance between the transducers is increased by x_3-x_1. The time wave form of the

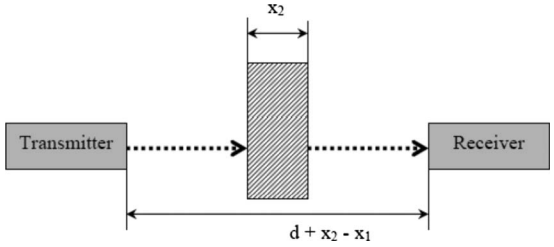


FIG. 8. Schematic of the attenuation coefficient measurement (x_2).

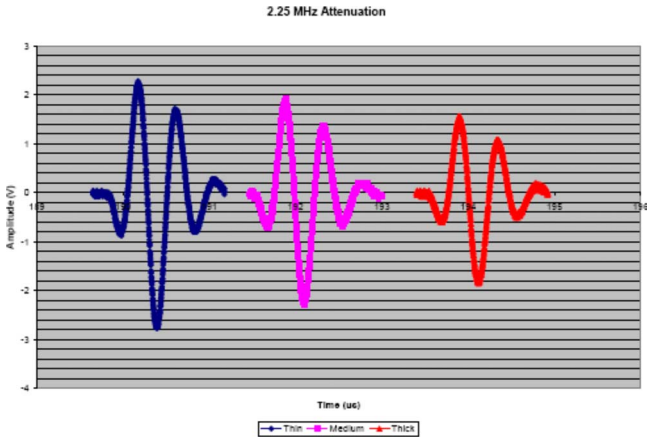


FIG. 9. (Color online) Collected data for estimation of the attenuation coefficient.

received signal is again stored in the computer and the attenuation coefficient is computed as

$$\alpha_f = \frac{20}{x_3 - x_1}(\log_{10} V_{\text{p.p.-}x_1} - \log_{10} V_{\text{p.p.-}x_3}) \text{ dB/cm}. \quad (6)$$

An example of the collected data at 2.25 MHz for three Plexiglas plates is presented in Fig. 9.

The average attenuation coefficient is computed for each operating frequency. Measurements are repeated with two other sets of transducers. The attenuation coefficients, in dB/cm per MHz, are computed by the following expression:

$$\alpha_{\text{dB/cm MHz}} = \frac{\alpha_{f_1} - \alpha_{f_2}}{f_1 - f_2}. \quad (7)$$

Attenuation of ultrasound waves can be used in the quality control of the materials, such as distance-amplitude calibration procedures.

E. NDE procedures

In order to incorporate the NDE procedures, the laboratory was equipped by the following devices: portable ultrasonic flaw detectors USN 58L and USM 35X

(Fig. 10), which allow the following experiments to be implemented:[8]

- calibration of automatic flaw detectors,
- evaluation of homogeneity of various materials used in industrial applications,
- detection and localization of heterogeneities in the materials, such as flaws, cavities, layers, and holes,[9–12] and

FIG. 10. (Color online) Portable ultrasonic flaw detectors USN 58L and USM 35X (from left to right).

FIG. 11. (Color online) Calibration of the flaw detectors using a straight-beam probe.

- measurement of the dimensions of various parts and components, where conventional methods (such as rulers and calipers) cannot be applied.

Students were engaged in the weekly experiments using the equipment described above. Specifically, the following experiments were carried out.

(1) Calibration of the flaw detectors using "Straight-Beam" single-element or dual-element probes utilizing the instruments' "AUTO-CAL" feature. During the procedure, students set the display "RANGE" so that two reference calibration echoes from different material thicknesses (Fig. 11) are displayed on the screen (Fig. 12).
The correct calibration is confirmed by the message "CALIBRATION IS DONE." The flaw detector then automatically determines the sound velocity of the material being tested in this procedure and the probe delay.
(2) Evaluation of the resolution of the system, i.e., the ability of the system to differentiate two or more discontinuities closely spaced in a lateral plane (Fig. 13). After the calibration of the instrument, the detection of discontinuities is carried out with the same instrument and the same probe.
(3) Calibration procedure with the "Angle-Beam" probe for wedge angle verification, sound path distance calibration, and flaw sensitivity calibration using an International Institute of Welding (IIW) type 1 calibration block and the results of the calibration are presented below (Fig. 14).

After calibration of the instrument, students place the transducer at various positions on the IIW block (Fig. 14) to display reflections from side drilled holes and evaluate sound path distances and sensitivity relationships. In addition, the students were able to detect and evaluate various defects in welds.

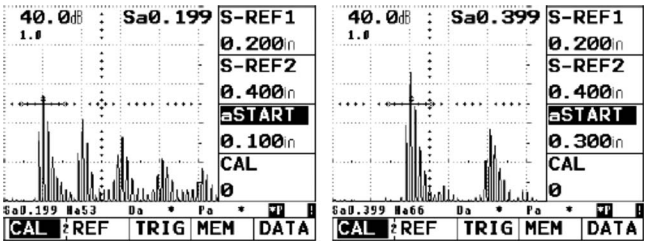


FIG. 12. First calibration echo (left) and second calibration echo (right).

FIG. 13. (Color online) Evaluation of the resolution of the system.

Upon completion of all laboratory sessions, each team was responsible for writing a final report that described the experimental techniques utilized, discussed the expected outcomes, provided data of the actual outcomes, and explained the reasons for the discrepancies between the expected and the actual results. Finally, the teams presented their findings to the class as a whole (Fig. 15).

IV. REMOTE CONTROL OF THE NDE EQUIPMENT

One of the main goals of the project is to develop a videoconference teaching NDE facility, which will provide greater program delivery flexibility and offer nontraditional educational segments.[13–15] This facility will allow all AET students at Drexel, as well as students at remote locations, to be involved in the same educational and training process in NDE. By expanding training opportunities to students who might not otherwise take advantage of them, due to distance and time, this facility helps reduce the shortage of trained specialists in NDE field. Key factors in the development process include creation of the educational laboratory that can significantly contribute to the development of technologically literate students and workforce that will be in great demand nationwide. The fully interactive videoconference teaching course in NDE was developed for undergraduate AET students and may also be taken by other undergraduate/graduate students at Drexel or by the students of other universities and community colleges who have fulfilled the necessary prerequisites and desire to pursue a BS degree in AET or obtain training in NDE. This approach will facilitate the development of teamwork that will allow the project/laboratory to proceed without the constant supervision of the faculty advisor.[16] The interinstitutional class sessions will be carried out utilizing Internet2-based access to the equipment of Drexel's NDE laboratory for other universities and community colleges. Students involved in the "live" interaction with other participants will share in small group discussions, collaborate, and fully engage in the videoconferencing expe-

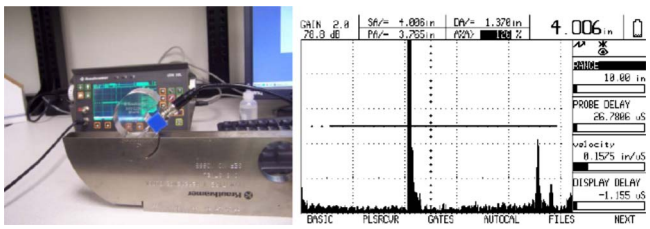


FIG. 14. (Color online) Calibration with the angle-beam probe for sound distance and flaw sensitivity using an HW calibration block.

FIG. 15. (Color online) Final presentation.

rience. The implications of fully interactive videoconference teaching are far reaching as they relate to distance delivery of real-time interactive instruction between any remote sites subscribing to Internet2 services using Internet Protocol networks. The state-of-the-art NDE facility is also designed to serve working individuals interested in improving their skills in NDE as well as those seeking knowledge for professional advancement.

Real-time remote control of USN 58L and USM 35X devices is under development. GE Inspection Technologies' ULTRADOC software allows for control and data transfer to and from the portable ultrasonic flaw detectors. Ultra virtual network computing (ULTRAVNC) software enables remote control and data transfer from the local computer connected to the flaw detectors and the camcorder simultaneously. Utilizing ULTRAVNC and ULTRADOC control function and commands, one can remotely control and change any setting of the flaw detectors, such as calibration of flaw detectors and evaluation of the test objects. This configuration of equipment and software packages allow students at the remote site for participation in the laboratory activities. Technician's or teaching assistant's presence at the local site is required for initial setup of the videoconferencing and NDE equipment and handling of the transducers. The calibration of the NDE equipment and testing procedures can be completely controlled from the remote site. The results of the calibration and testing can be saved from both local and remote computers. The block diagram of the remote NDE procedure is presented in Fig. 16.

Several functions using LABVIEW controller for the remote NDE procedures have already been developed. An example of the remote calibration procedure using the LABVIEW VI controller is presented in Fig. 17.

V. COURSE EVALUATION

The described course has been designed with educational objectives and learning outcomes based on the general Accreditation Board for Engineering Technology (ABET) criteria for engineering technology programs. Multiple assessment measures have been implemented and documented to demonstrate that the course objectives and outcomes are being met. A description of the assessment and evaluation is provided below.

The EET-203 Nondestructive Evaluation of Materials course was offered during the fall term of 2007–2008 academic year. Evaluation of the course and data collection began soon after the start of the term. Students who enrolled in EET-203 were administered a pretest, which assessed the entering knowledge requirements for the course. Students' knowledge of appropriate algebra, trigonometry, and vibrations and sound waves were evaluated. Based on the results of the test, students were divided into four groups according to the "think-share-report-learn" process,[17] which involved peer coaching to help each other during the laboratory procedures. The effectiveness of this approach to teaching the problem-based hands-on learning model within a virtual environment of a state-of-the-art laboratory for NDE purposes was evaluated using both formative and summative evaluations.[17] A formative evaluation assessed initial and ongoing project activities based on the students' tests, laboratory reports (which were required within seven days after each laboratory session), and class discussions. A summative evaluation assessed the quality and impact of an implemented project based on the students' final presentations, including corrections of the collected results and conclusions,

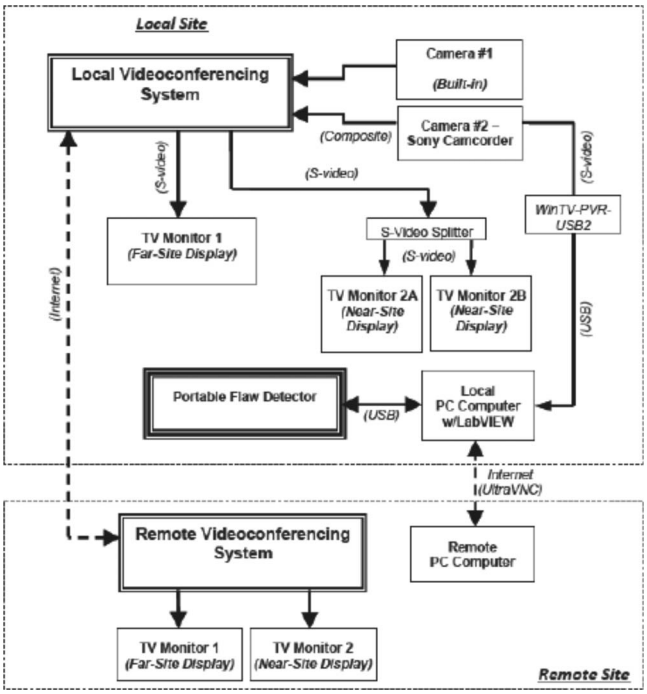


FIG. 16. Block diagram of the remote NDE procedure.

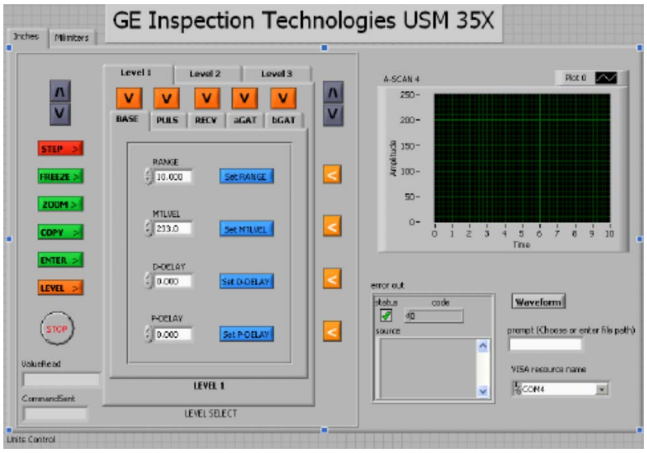


FIG. 17. (Color online) LABVIEW controller of the NDE calibration procedure.

TABLE I. Likert-type scale.

As an AET engineer, do you feel confident in NDE?[a]			
Not at all confident	A little confident	Pretty confident	Very confident
As an AET engineer, do you feel knowledgeable in NDE?			
Not al all knowledgeable	A little knowledgeable	Pretty knowledgeable	Very knowledgeable
As an AET engineer, do you feel competent in NDE?			
Not at all competent	A little competent	Pretty competent	Very competent

[a]Put an "X" on the line under the correct answer that shows how you feel about each statement.

and course evaluations. Industrial evaluators were involved in both formative and summative evaluations. Based on their recommendations, the laboratory procedures were modified according to the requirements of ASNT. After completion of all laboratory procedures, students completed a course evaluation form, including the following questions according to a Likert-type scale (Table I), which assessed the students' perception of their confidence, knowledge, and competence.

VI. SUMMARY

During the fall term of the 2007–2008 academic year, the developed course EET-203 (Nondestructive Evaluation of Materials) was offered to prejunior and junior AET students. The industrial partners, in collaboration with the faculty, developed real-world industrial problems. Mr. Tony Midora of Precision Measurement Technology Inc. and Mr. Richard Bottari of Stork MMA Testing Laboratories participated in providing instruction during the laboratory sessions.

Based on the feedback from the industrial collaborators and students' course evaluations, the necessary changes were made in laboratory procedures. The industrial collaborators also provided samples of parts and materials used for inspection. After completion of all laboratory sessions, the students became familiar with basic acoustical instrumentation, possessed hands-on experience with ultrasonic and electronic equipment, and were able to demonstrate the basic principles of ultrasound measurements and NDE techniques. An important objective of this laboratory-based course was to improve the students' knowledge of data gathering, the identification of sources leading to erroneous measurements, and proficiency in communication skills. Therefore, a concise written report clearly describing all conclusions and comments was required within seven days after completion of each laboratory session. Students worked in teams on projects drawn from several areas of technological interest.

The simulation of the NDE applications used by companies in industry was implemented in the proposed project. Qualified evaluators from industry made an evaluation of the success of the course based on the students' laboratory reports, the final report, and the final presentation. This project is still under development. Additional NDE equipment will be purchased and installed. During the development of this project, more experiments and case studies will be added to this course.

ACKNOWLEDGMENT

This work is supported by NSF Grant No. 0632734 "Implementation of the Internet-Based Nondestructive Evaluation Laboratory for Applied Engineering Technology Curriculum."

[1]D. J. Hagemaier, "Aerospace applications of ultrasonic testing," in *American Society for Nondestructive Testing (ASNT) Handbook* (ASNT, Columbus, OH, 1991), Vol. **7**, Chap. 14, pp. 493–515.

[2]M. A. Breazeale, J. H. Cantrell, Jr., and J. S. Heyman, in *Methods of Experimental Physics*, Ultrasonics Vol. **19**, edited by P. H. Edmonds (Academic, New York, 1981), pp. 67–135.

[3]E. P. Papadakis, "The measurement of ultrasonic velocity and the measurements of ultrasonic attenuation," in *Ultrasonic Measurement Methods*, Physical Acoustics Vol. **XIX**, edited by R. N. Thurston and A. D. Pierce (Academic, Boston, 1990), pp. 81–106.

[4]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000), Chap. 7, pp. 171–209.

[5]E. P. Papadakis, "Theoretical and experimental methods to evaluate ultrasonic transducers for inspection and diagnostic applications," IEEE Trans. Sonics Ultrason. **SU-26**, 14–25 (1979).

[6]E. P. Papadakis, "Ultrasonic attenuation and velocity in three transformation products in steel," J. Appl. Phys. **35**, 1474–1482 (1964).

[7]E. P. Papadakis, "The measurement of ultrasonic velocity and the measurements of ultrasonic attenuation," in *Ultrasonic Measurement Methods*, Physical Acoustics Vol. **XIX**, edited by R. N. Thurston and A. D. Pierce (Academic, Boston, MA, 1990), pp. 107–155.

[8]D. Fei, D. Hsu, and M. Warchol, "Simultaneous velocity, thickness and profile imaging by ultrasonic scan," J. Nondestruct. Eval. **20**, 95–112 (2002).

[9]*Basic Principles*, Nondestructive Testing Vol. **1** (General Dynamics, American Society for Nondestructive Testing, 2002).

[10]*Equipment*, Nondestructive Testing Vol. **II** (General Dynamics, American Society for Nondestructive Testing, 2002).

[11]*Applications*, Nondestructive Testing Vol. **III** (General Dynamics, American Society for Nondestructive Testing, 2002).

[12]P. Mix, *Introduction to Nondestructive Testing. A Training Guide* (Wiley, Hoboken, NJ, 2005), pp. 457–520.

[13]B. Aktan, C. A. Bohus, L. A. Crowl, and M. H. Shor, "Distance learning applied to control engineering laboratories," IEEE Trans. Educ. **39**, 320–326 (1996).

[14]M. Casini, D. Prattichizzo, and A. Vicino, "The automatic control telelab: A user-friendly interface for distance learning," IEEE Trans. Educ. **46**, 252–257 (2003).

[15]H. Shen, Z. Xu, B. Dalager, V. Kristiansen, O. Strom, M. S. Shur, T. A. Fjeldly, J.-Q. Lu, and T. Ytterdal, "Conducting laboratory experiments over the Internet," IEEE Trans. Educ. **42**, 180–185 (1999).

[16]V. Genis, A. Daneshpooy, and C. O'Shaughnessy, "Development of videoconference teaching for applied engineering technology students," Proceedings of the American Society for Engineering Education (ASEE) Mid-Atlantic Conference, April (2005), pp. 1–6.

[17]J. F. Westat, *The 2002 User Friendly Handbook for Project Evaluation (NSF 02-057)* (The National Science Foundation, Arlington, VA, 2000).

Effective medium method for sound propagation in a soft medium containing air bubbles

Bin Liang, Xinye Zou, and Jianchun Cheng[a)]
Laboratory of Modern Acoustics and Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

(Received 4 December 2007; revised 11 June 2008; accepted 16 June 2008)

An effective medium method (EMM) is developed to investigate the nonlinear propagation of acoustic waves for soft media containing air bubbles, which accounts for the effects of weak compressibility, viscosity, surrounding pressure, surface tension, and encapsulating shells. Based on the dynamics model of an individual bubble that has included these effects, the EMM is presented by employing a simple perturbation approach to "homogenize" the bubbly soft media. The equations describing the fundamental and the second harmonic waves are derived that applies to three-dimensional cases, and then solved in a one-dimensional case to obtain the effective acoustical parameters of a longitudinal wave. The EMM is compared with the previous theories in three representative cases regarded as simple models of significant practical applications. The results show that the EMM agrees well with the previous theories and can incorporate the additional effects, which may notably affect the accuracy of the results. The limitations of the EMM are also identified and stated. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2957931]

PACS number(s): 43.25.Dc, 43.20.Hq, 43.25.Yw [YHB] Pages: 1419–1429

I. INTRODUCTION

Attempts to theoretically investigate the bubble dynamics in a viscoelastic medium go back many decades,[1,2] and research in this field is still active due to the potential applications.[3–6] Meyer *et al.* have performed the early measurements of the resonance frequency of a bubble.[1] The dynamical equation for arbitrarily large radial motion has been derived by Eringen and Suhubi[2] in Rayleigh–Plesset form. It has been proved, however, that only if the ratio λ/μ is sufficiently large can a bubble in a solid behave like an oscillator with a large quality factor.[7] Here λ and μ are the Lamé coefficients. In this situation, it is important to study the bubble dynamics in a soft medium, e.g., a tissue or rubberlike medium, for which the inequality $\lambda \gg \mu$ is satisfied. Ostrovsky[3,4] has derived a Rayleigh–Plesset-like equation to describe the nonlinear oscillation of an individual bubble in a soft medium, but only in the unstressed equilibrium state, which occurs when the equilibrium gas pressure in the bubble equals the pressure at infinity. Emelianov *et al.*[5] have studied the nonlinear dynamics of a bubble in an incompressible medium in the nonuniformly equilibrium state that occurs when these two pressures are different. The model of Emelianov *et al.* has then been extended by Zabolotskaya *et al.*[6] to include the effects of surface tension, viscosity, weak compressibility, and confinement by an elastic shell.

An air bubble surrounded by a soft medium is the elementary structural unit of a soft medium permeated with bubbles that applies to a variety of significant engineering situations, e.g., the biomedical applications utilizing microbubbles.[6] As a matter of fact, the propagation of an acoustic wave in a soft medium permeated with air bubbles will be remarkably affected by the existence of the bubbles. Ostrovsky[3,4] has proved that the nonlinear nature of the bubble oscillation can strongly enhance the nonlinearity of a bubbly soft medium. In practical situations, the acoustical properties of a bubbly soft medium are notably influenced by other factors that are associated with bubble oscillations as well, such as surface tension, viscosity, compressibility, surrounding pressure, and an encapsulating elastic shell. Consequently, there are particularly significant practical reasons for developing a theoretical description of the nonlinear acoustical properties of a bubbly soft medium with these effects included. Considerable efforts have been, and continue to be, devoted to the theoretical study of propagation of acoustic wave in inhomogeneous media and many theories have been presented.[8–16] Most of the previous theories are developed under linear approximation and account only for the linear acoustical properties of inhomogeneous media.[8–13] Ma *et al.*[14] have studied the nonlinearity of liquid containing encapsulated bubbles. The nonlinearity of a bubbly soft medium has been analyzed by Liang and Cheng[15] to include the effect of viscosity, and later by Qin *et al.*[16] in nonuniformly equilibrium states. To our knowledge, however, a suitable theory is still to be proposed to describe precisely the nonlinear acoustical properties of bubbly soft media with the incorporation of the effects of surface tension, viscosity, surrounding pressure, and encapsulating shells.

The principal purpose of the present article is to provide formalism for describing the propagation of an acoustic wave in a soft medium containing air bubbles, and to estimate the influence of the effects discussed earlier on the linear and nonlinear acoustical properties of the bubbly soft medium. In the present study, we employ the homogeneous approximation on the basis of the bubble dynamics model presented Zabolotskaya *et al.*[6] for such a bubbly soft medium, and thereby develop an effective medium method (EMM) to give a description of the nonlinear sound propa-

[a)]Electronic mail: jccheng@nju.edu.cn

gation in the bubbly medium with these effects accurately included. Note that many previous methodologies investigate the wave propagation in inhomogeneities by treating the media as a homogeneous "effective" medium characterized by effective parameters as well. Despite the similarity between the EMM and these methodologies, the present study definitely differs from the previous ones in several respects. It is a fundamental distinction that the EMM accounts for the nonlinearity of the bubbly soft medium up to a second-order approximation, whereas most of the previous ones only use linear approximation when homogenizing the medium. For such a medium with particularly strong "physical" nonlinearity, which stems from inhomogeneous structures,[3,4] linearizing the fundamentally nonlinear wave equation inevitably loses significant details. Second, the EMM permits one to take into consideration the effects of weak compressibility, surface tension, viscosity, surrounding pressure, and an encapsulating elastic shell. There are important practical reasons for pursuing more precise results in various engineering situations, for which the incorporation of these effects is apparently necessary. One such example is the clinic application of ultrasonic imaging utilizing the contrast agents, which will be employed as a particular case of numerical experiment and studied later in detail. Note that the actual values of human blood pressures may vary within a considerably large range when compared with the atmospheric pressure. As will be proved in this case, the effect of ambient pressure has to be properly estimated to achieve an accurate result. Finally, the EMM could apply to three-dimensional cases rather than one-dimensional cases. Most of the relative studies investigate only the wave propagation in an infinite effective medium for which the one-dimensional approximation is sufficient, but it is indispensable to obtain the three-dimensional effective parameters for some practical structures of finite sizes. It must be stressed, however, that there also exist limitations of the application of the EMM despite its effectiveness, and the validity of the EMM is thus by no means unconditional. First, the EMM is developed under quadratic approximation by employing a simple perturbation approach, and the nonlinearity of medium is studied by inspecting the second harmonic wave with no harmonics of orders higher than 2 involved. Second, the present model could not enable full incorporation of all the practical effects that affect the acoustical properties of a bubbly medium. For a typical instance, the bubbles are assumed to approximately remain spherical shapes in the present model. Then, the EMM will inevitably fail to give accurate results in some special cases where the curvatures of the surfaces of bubbles undergo irregular distortions, such as the cases where the buckling of bubbles dominates and the effective nonlinearity parameters thus take large and negative values.[17] These problems will be the focus of a future study.

The present article is organized as follows. In Sec. II, we start from the model of Zabolotskaya *et al.*[6] regarding the dynamics of an individual bubble in a soft medium with the effects (compressibility, surface tension, etc.) incorporated, for which the analysis is performed in terms of bubble radius. The bubble dynamics equation presented in their work is rewritten into another form that is expressed in terms of bubble volume variation, for the purpose of facilitating the present study. On the basis of this oscillation equation, an EMM is developed in Sec. III for describing the propagation of an acoustic wave in a soft viscoelastic medium permeated with encapsulated bubbles. By using a perturbation approach up to second-order approximation, the wave equations are derived for the linear and nonlinear wave propagation. Effective acoustical parameters of a longitudinal wave are obtained by solving the wave equations in a simple one-dimensional case. In Sec. IV, the applications of the EMM as well as some previous methods are analyzed numerically in three particular cases that may be the simple model of significant practical situations, as follows: (1) blood with contrast agents injected; (2) a soft medium containing contrast agents; and (3) a soft medium permeated with free bubbles. The comparison between the EMM and the other ones prove that the EMM is valid and can improve the accuracy of the results when the influences of the additional effects on the effective acoustical parameters of the bubbly media are nontrivial.

II. BUBBLE DYNAMICS

Consider an encapsulated gas bubble surrounded by a soft viscoelastic medium. When in equilibrium, the gas pressure in the bubble is denoted P_g, and the pressure infinitely far away is P_∞. For the case where the equilibrium pressure equals the surrounding pressure (i.e., $P_g=P_\infty$), the shear stress is uniform throughout the soft medium. Such a case is referred to as an initially unstressed state, for which the equilibrium values of the inner and outer radius of the bubble are designated R_0 and R_{s_0}, respectively. In the general case, however, the encapsulated bubble may be pressurized, such that $P_g \neq P_\infty$. Such a case is denoted as a prestressed case due to the fact that a nonuniform shear stress is generated inside the medium to balance the pressure difference. For prestressed cases we define the equilibrium values of the inner and outer radii as R_1 and R_{s_1}, respectively. The geometry is shown in Fig. 1. Figure 1(a) shows an unstressed case where one has $R_0=R_1$ and $R_{s_0}=R_{s1}$. In the cases where $P_g<P_\infty$, however, it is apparent that the pressure difference between P_g and P_∞ will force the bubble to shrink, and one thus has $R_0>R_1$ and $R_{s_0}>R_{s_1}$, as illustrated in Fig. 1(b). In contrast, one has $R_0<R_1$ and $R_{s_0}<R_{s_1}$ if $P_g>P_\infty$. As the bubble oscillates, the instantaneous values of the inner and outer radii are defined as $R(t)$ and $R_s(t)$, respectively.

Zabolotskaya *et al.*[6] have studied the nonlinear dynamics in the form of a Rayleigh–Plesset-like equation for an individual bubble in such a model, provided the approaches include the effects of compressibility, surface tension, viscosity, and an encapsulating shell. Note that Eq. (53) in Ref. 6 accounts for the effects of surface tension, viscosity, and shell, but applies only to the case of an incompressible medium. Adding the compressibility term $(d^3w/dt^3)/c_m$ that accounts for the radiation damping to the left-hand side of this equation, one readily obtains the equation that describes the nonlinear oscillation of a single bubble in a soft medium, as follows:

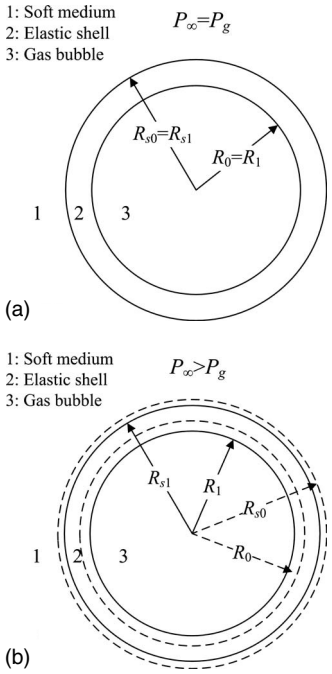


FIG. 1. Geometry of an encapsulated gas bubble in a soft medium in (a) an initially unstressed state and (b) a prestressed state.

$$F(R)R\frac{d^2R}{dt^2} + \frac{3}{2}G(R)\left(\frac{dR}{dt}\right)^2 - \frac{1}{c_m}\frac{d^3w}{dt^3} = \frac{1}{\rho_m}\left\{P_g\left(\frac{R_1}{R}\right)^{3\gamma} - P_\infty - P_e(R) - \frac{2\sigma_g}{R} - \frac{2\sigma_m}{R_s} - \frac{4}{R}\frac{dR}{dt} \times \left[\eta_s\left(1 - \frac{R^3}{R_s^3}\right) + \eta_m\frac{R^3}{R_s^3}\right]\right\}. \quad (1)$$

In the preceding expression, the parameter w is defined as $w=R^3/3$, $P_\sigma=2\sigma_g/R+2\sigma_m/R_s$ is the effective pressure due to the surface tension with σ_g and σ_m being the surface tensions at the inner gas–shell interface and the outer shell–medium interface, respectively. ψ is the dissipation function that is found to be $\psi=8\pi R(dR/dt)^2[\eta_s(1-R^3/R_s^3) + \eta_m R^3/R_s^3]$ with η_s and η_m being the shear viscosity coefficients in the shell region and in the medium, respectively. The parameters of $F(R)$ and $G(R)$ are given as

$$F(R) = \frac{\rho_s}{\rho_m} + \left(1 - \frac{\rho_s}{\rho_m}\right)\frac{R}{R_s},$$

$$G(R) = \frac{\rho_s}{\rho_m} + \left(1 - \frac{\rho_s}{\rho_m}\right)\left(\frac{4}{3} - \frac{1}{3}\frac{R^3}{R_s^3}\right)\frac{R}{R_s},$$

where ρ_s and ρ_m are the mass densities of the shell and the surrounding medium, respectively, and $P_e(R)$ refers to the effective pressure due to the strain energy stored in shear deformation of both the shell and the medium, defined as[6]

$$P_e(R) = 4\int_r^{R_s}\left(\frac{\partial\varepsilon_s}{\partial I_1} + \frac{r^2}{\tilde{r}^2}\frac{\partial\varepsilon_s}{\partial I_2}\right)\left(\frac{r^2}{\tilde{r}^2} - \frac{\tilde{r}^4}{r^4}\right)\frac{dr}{r} + 4 \times \int_{R_s}^{\infty}\left(\frac{\partial\varepsilon_m}{\partial I_1} + \frac{r^2}{\tilde{r}^2}\frac{\partial\varepsilon_m}{\partial I_2}\right)\left(\frac{r^2}{\tilde{r}^2} - \frac{\tilde{r}^4}{r^4}\right)\frac{dr}{r}, \quad (2)$$

where $\varepsilon=\varepsilon(I_1,I_2,I_3)$ refers to the strain energy density with I_1, I_2, and I_3 being the principal invariants of Green's deformation tensor, r and $\tilde{r}$ refer to the Eulerian and the Lagrangian coordinates, respectively, and the subscripts s and m refer to the shell and the surrounding medium, respectively.

For the convenience of the following investigation, Eq. (1) will be evaluated here in the quadratic approximation by rewriting it into another form for the perturbation in bubble volume defined as $U=4\pi(R^3-R_1^3)/3$. For a soft medium, Mooney's constitutive relation[18] is the most widely used model equation and has been adopted by many previous studies regarding the nonlinear dynamics of a bubble in such a medium.[3–6] For facilitating the comparison with the previous studies, therefore, we employ Mooney's relation to evaluate the effective pressure $P_e(R)$, as follows:

$$\varepsilon_p = \mu_p[(1+\chi)(I_1-3) + (1-\chi)(I_2-3)]/4, \quad p = s,m, \quad (3)$$

where μ_p is the shear modulus and χ is the fitting parameter referred to by Mooney as the coefficient of asymmetry that must lie in the range $-1\leqslant\chi\leqslant 1$.

Substituting Eq. (3) into Eq. (2) and expanding $P_e(R)$ to quadratic order, one may derive an analytical approximation of $P_e(R)$, as follows:

$$P_e(R) = P_e(\zeta) = P_e(\zeta_1) + (\zeta-\zeta_1)P_e'(\zeta_1) + (\zeta-\zeta_1)^2P_e''(\zeta_1)/2,$$

where $\zeta=R/R_0$, $\zeta_1=R_1/R_0$, the primes represent derivatives with respect to ζ, and the parameters of $P_e(\zeta)$, $P_e'(\zeta)$, and $P_e''(\zeta)$ are given as

$$P_e(\zeta) = \mu_m\int_0^a[(1+\chi)(yx^{-1}-y^7x^{-8}) + (1-\chi) \times (y^{-1}-y^5x^{-6})]dx + \mu_s\int_a^1[(1+\chi)(yx^{-1}-y^7x^{-8}) + (1-\chi)(y^{-1}-y^5x^{-6})]dx, \quad (4a)$$

$$P_e'(\zeta) = \mu_s\zeta^2\int_0^a[(1+\chi)(7y^{10}x^{-8}-y^4x^{-2}) + (1-\chi)(y^2 + 5y^8x^{-6})]dx + \mu_m\zeta^2\int_a^1[(1+\chi)(7y^{10}x^{-8}-y^4x^{-2}) + (1-\chi)(y^2+5y^8x^{-6})]dx, \quad (4b)$$

$$P_e''(\zeta) = \mu_s \zeta^4 \int_0^a [(1+\chi)(4y^7x^{-2} - 70y^{13}x^{-8}) - (1-\chi) \times (2y^5 + 40y^{11}x^{-6})]dx + \mu_m \zeta^4 \int_a^1 [(1+\chi)(4y^7x^{-2} - 70y^{13}x^{-8}) - (1-\chi)(2y^5 + 40y^{11}x^{-6})]dx + 2P_e'(\zeta)/\zeta, \quad (4c)$$

where $x=R_0/\tilde{r}$, $y=R_0/r$, and $a=R_0/R_{s_0}$.

Substituting Eq. (4) into Eq. (1), one obtains the expansion of Eq. (1) to quadratic order in U, as follows:

$$\frac{d^2U}{dt^2} + \delta_1 \frac{dU}{dt} + \omega_1^2 U - \frac{R_1}{F_1 c_m}\frac{d^3U}{dt^3} = G_1 U^2 + \delta_2 U \frac{dU}{dt} + H_1\left[\left(\frac{dU}{dt}\right)^2 + 2U\frac{d^2U}{dt^2}\right] - e_1 P_A, \quad (5)$$

where $P_A(t)=P_\infty - P_0$ is the applied acoustic pressure with P_0 being the pressure at infinity in the absence of sound and $\omega_1=\sqrt{\omega_g^2+\omega_e^2-\omega_\sigma^2}$ is the nature frequency of bubble for which the components are given as

$$\omega_g^2 = 3\gamma P_g D_1, \quad \omega_e^2 = P_e'(\zeta_1)\zeta_1 D_1,$$

$$\omega_\sigma^2 = 2D_1 R_1^{-1}(\sigma_g + \sigma_m \gamma_R^4),$$

where γ is the ratio of specific heats, which is chosen as $\gamma = 1.4$ as only air bubbles are considered in the present article, δ_1 and δ_2 are the viscous damping coefficients at linear and quadratic order, respectively, defined as

$$\delta_1 = 4D_1[\eta_s(1-\gamma_R^3) + \eta_m \gamma_R^3],$$

$$\delta_2 = 24 q_1 D_1[\eta_s(1-\gamma_R^6) + \eta_m \gamma_R^6],$$

G_1 is the nonlinearity coefficient associated with gas compressibility, elasticity, and surface tension that is defined as

$$G_1 = q_1\left[3(\gamma+1)\omega_g^2 + \left(2 - \frac{P_e''(\zeta_1)\zeta_1}{P_e'(\zeta_1)}\right)\omega_e^2 - 4\omega_\sigma^2 + \frac{8\sigma_m R_1}{\rho_m F_1 R_{s1}^4}(1-\gamma_R^3)\right],$$

and the parameters of F_1, H_1, e_1, q_1, γ_ρ, γ_R, and D_1 are given as follows:

$$F_1 = \gamma_\rho + (1-\gamma_\rho)\gamma_R, \quad H_1 = q_1 F_1^{-1}[\gamma_\rho + (1-\gamma_\rho)\gamma_\rho^4],$$

$$e_1 = 4\pi D_1 R_1^3, \quad q_1 = 1/(8\pi R_1^3),$$

$$\gamma_\rho = \rho_s/\rho_m, \quad \gamma_R = R_1/R_{s1}, \quad \text{and } D_1 = 1/(\rho_m F_1 R_1^2).$$

III. EFFECTIVE MEDIUM METHOD

A. Effective medium

We now study the propagation of a plane acoustic wave in an infinite soft medium permeated with random encapsulated bubbles, subject to the condition that the volume content of the bubbles is small but the number of bubbles on a scale of wavelength order is large. Then it can be proved that the multiple scattering effects are negligible[8] and the homogeneous approximation, well known for liquid containing bubbles, can be employed.[3] Consider a small volume element of medium length dx_i in the x_i direction in the Cartesian coordinate $(i=1,2,3)$ that is sufficiently large to include a number of bubbles. In the present study, we shall focus our attention on the cases where the amplitude of a wave is small, for the purpose of investigating the strong physical nonlinearity of such a class of media.[3,4] Then, the dynamic nonlinearity is negligible, which dominates only on the condition that the amplitude of wave is finite. According to the stress–strain relationship and neglecting the contribution of the gas inside the bubbles, the stress tensor may be expressed as (see Ref. 19, p. 10)

$$\sigma_{ik}^s = (1-\beta)[K u_{ll}^s \delta_{ik} + 2\mu(u_{ik}^s - \delta_{ik}u_{ll}^s/3)], \quad (6)$$

where σ_{ik}^s and u_{ik}^s are the stress tensor and the strain tensor of the solid phase, respectively, β is the total volume fraction of the bubbles, $K=\lambda+2\mu/3$ is the bulk modulus, and δ_{ik} refers to the Kronecker delta which is defined as

$$\delta_{ik} = \begin{cases} 1, & i=k \\ 0, & i \neq k \end{cases}.$$

On the other hand, the volume element of the bubbly soft medium may be regarded as a volume element of effective medium that is homogeneous and is described by effective acoustical parameters. The stress tensor of the effective medium may be given as

$$\tilde{\sigma}_{ik} = \tilde{K}\tilde{u}_{ll}\delta_{ik} + 2\tilde{\mu}(\tilde{u}_{ik} - \delta_{ik}\tilde{u}_{ll}/3), \quad (7)$$

where $\tilde{\sigma}_{ik}$ and $\tilde{u}_{ik}$ are the stress tensor and the strain tensor of the effective medium, respectively, and $\tilde{K}=\tilde{\lambda}+2\tilde{\mu}/3$ is the effective bulk modulus with $\tilde{\lambda}$ and $\tilde{\mu}$ being the effective Lamé coefficients of the effective medium.

B. Influence of bubble oscillation

As the acoustic wave propagates in the bubbly medium, the volume of the bubbles will change due to the oscillation of the bubbles driven by the acoustic wave. As a result, the variation of the volume element includes the compression of the elastic phase and the variation of the total volume of the bubbles. Then, one has

$$\tilde{\theta} = \theta^s(1-\beta) + V_t, \quad (8)$$

where $\tilde{\theta}=\tilde{u}_{11}+\tilde{u}_{22}+\tilde{u}_{33}$ and $\theta^s=u_{11}^s+u_{22}^s+u_{33}^s$ are the volume changes of the effective medium and the elastic phase, respectively, and V_t is the variation of the specific volume of bubbles.

As the bubble distorts under the action of the shear deformation, the principal radii of curvature of the surface will change. This will certainly change the effect of surface tension and then the bubble volume. In the present study, however, the bubbles are assumed spherical, and such an effect is then negligible that does not change the nature of the bubble dynamics. Then it is fair to assume approximately that the pure shear deformation of the volume element will not be affected by the existence of bubbles. Then one has

$$\sigma_{ik}^s = \tilde{\sigma}_{ik}, u_{ik}^s = \tilde{u}_{ik}, \quad i \neq k. \tag{9}$$

Substituting Eq. (9) into Eqs. (6) and (7) yields $\tilde{\mu}=\mu(1-\beta)$. From Eqs. (6) and (7) one readily obtains

$$\sigma_{11}^s + \sigma_{22}^s + \sigma_{33}^s = 3(1-\beta)(\lambda + 2\mu/3)\theta^s, \tag{10a}$$

$$\tilde{\sigma}_{11} + \tilde{\sigma}_{22} + \tilde{\sigma}_{33} = 3(\tilde{\lambda} + 2\tilde{\mu}/3)\tilde{\theta}. \tag{10b}$$

Under the action of an applied force, the element of effective medium is defined to produce the same stress as the element of the bubbly medium. Hence one has

$$\sigma_{11}^s + \sigma_{22}^s + \sigma_{33}^s = \tilde{\sigma}_{11} + \tilde{\sigma}_{22} + \tilde{\sigma}_{33}. \tag{11}$$

Substituting Eqs. (8)–(10) into Eqs. (6) and (7) yields (for $\lambda \gg \mu$)

$$\tilde{C}_{11}\theta^s = C_{11}(\theta^s - V_t), \tag{12}$$

where $C_{11}=\lambda+2\mu$, $\tilde{C}_{11}=\tilde{\lambda}+2\tilde{\mu}$ are the elastic modulus of the soft medium and the effective medium, respectively.

For the purpose of solving the unknown quantity V_t, it is necessary to obtain the variation of the volume of an individual bubble U, which is described by the equation for the oscillation of a bubble given by Eq. (5). In Eq. (5), the acoustic pressure P_A accounts for the driving force of the oscillation of the bubble. Due to the fact the shear wave does not change the volume of the bubble, the driving force of the bubble oscillation is not affected by the shear wave but determined by the total radial force exerted by the incident wave.[3] According to Eq. (17) in Ref. 3 one has

$$P_A = \frac{\lambda + 2\mu/3}{\lambda + 2\mu}\sigma \simeq \sigma, \tag{13}$$

where $\sigma=\tilde{\sigma}_{11}$ is the pressure generated by the incident wave.

Substituting Eq. (13) into Eq. (5) yields

$$\frac{d^2U}{dt^2} + \delta_1\frac{dU}{dt} + \omega_1^2 U - \frac{R_1}{F_1 c_m}\frac{d^3U}{dt^3} = G_1U^2 + \delta_2 U\frac{dU}{dt} + H_1\left[\left(\frac{dU}{dt}\right)^2 + 2U\frac{d^2U}{dt^2}\right] - e_1\sigma. \tag{14}$$

Owing to the fact that Eq. (14) is nonlinear only to second order, a potential solution has the form

$$U \approx U_1 \exp(i\omega t) + U_2 \exp(i2\omega t) + \text{c.c.}, \tag{15a}$$

$$\sigma \approx \sigma_1(\mathbf{r})\exp(i\omega t) + \sigma_2(\mathbf{r})\exp(i2\omega t) + \text{c.c.}, \tag{15b}$$

where σ_1 and σ_2 refer to the linear and the nonlinear waves, respectively, U_1 and U_2 refer to the amplitude of the linear pulsation and the nonlinear response, respectively, $\mathbf{r}$ refers to the three-dimensional space coordinate position of the field point that may be expressed in the Cartesian coordinate as $\mathbf{r}=x_1\hat{i}+x_2\hat{j}+x_3\hat{k}$.

Substituting Eq. (15) into Eq. (14) and assuming that $1 > U_1 > U_2$ yields

$$U_1 = g_1\sigma_1, \quad U_2 = g_2\sigma_2 + \Gamma\sigma_1^2, \tag{16}$$

where

$$g_1 = -\frac{e_1}{(\omega_1^2 - \omega^2) + i(\omega\delta_1 + \omega^3 R_1/c_m)},$$

$$g_2 = -\frac{e_1}{(\omega_1^2 - 4\omega^2) + i(2\omega\delta_1 + 8\omega^3 R_1/c_m)},$$

$$\Gamma = g_1^2\frac{(G_1 - 3\omega^2 H_1) + i\omega\delta_2}{(\omega_1^2 - 4\omega^2) + i(2\omega\delta_1 + 8\omega^3 R_1/c_m)}.$$

Expanding the displacement vector $\mathbf{u}$, up to second order approximation, as

$$\mathbf{u} = \mathbf{u}_1(\mathbf{r})\exp(i\omega t) + \mathbf{u}_2(\mathbf{r})\exp(i2\omega t) + \text{c.c.},$$

one obtains

$$\sigma_1 = \tilde{C}_{11}^{(1)}\nabla\cdot\mathbf{u}_1, \quad \sigma_2 = \tilde{C}_{11}^{(2)}\nabla\cdot\mathbf{u}_2. \tag{17}$$

If the size distribution function of the bubbles is specified as $n(R_0)$ [so that $n(R_0)dR_0$ is the number of bubbles with radii from R_0 to R_0+dR_0 in unit volume], the variation of the specific volume of the bubbles V_t is related to the volume variation of an individual bubble U by the relationship

$$V_t = \int Un(R_0)dR_0. \tag{18}$$

Expanding V_t to a second-order approximation as

$$V_t = V_1\exp(i\omega t) + V_2\exp(i2\omega t) + \text{c.c.}, \tag{19}$$

one obtains from Eqs. (16)–(19)

$$V_1 = V_{g1}\tilde{C}_{11}^{(1)}\nabla\cdot\mathbf{u}_1, \tag{20a}$$

$$V_2 = V_{g2}\tilde{C}_{11}^{(2)}\nabla\cdot\mathbf{u}_2 + V_\Gamma(\tilde{C}_{11}^{(1)}\nabla\cdot\mathbf{u}_1)^2, \tag{20b}$$

where

$$V_{g1} = \int g_1 n(R_0)dR_0, \quad V_{g2} = \int g_2 n(R_0)dR_0,$$

$$V_\Gamma = \int \Gamma n(R_0)dR_0.$$

If all the bubbles are of the uniform radius R_0, V_t is related to U by the relationship $V_t=NU$ with $N=3\beta(4\pi R_0^3)^{-1}$ being the number of bubbles in unit volume. In such cases one has

$$V_{g1} = Ng_1, \quad V_{g2} = Ng_2, \quad V_\Gamma = N\Gamma.$$

Expressing the volume change of the solid phase as $\theta^s=\nabla\cdot\mathbf{u}$, one may rewrite Eq. (12) as follows:

$$\tilde{C}_{11}^{(1)}\nabla\cdot\mathbf{u}_1 = C_{11}(\nabla\cdot\mathbf{u}_1 - V_1), \tag{21a}$$

$$\tilde{C}_{11}^{(2)}\nabla\cdot\mathbf{u}_2 = C_{11}(\nabla\cdot\mathbf{u}_2 - V_2). \tag{21b}$$

C. Wave equations

According to Ref. 6, $\rho_m F_1$ is defined as the effective density of the soft medium surrounding the bubbles, the effective density of the effective medium may thus be identi-

fied as $\tilde{\rho}=\rho_m F_1(1-\beta)+\rho_g\beta$. As the dynamic nonlinearity of the medium associated with the finite amplitude of wave has been ignored, the wave equation of the effective medium may be written as follows:

$$\nabla(\tilde{C}_{11}\boldsymbol{\nabla}\cdot\mathbf{u})-\tilde{\mu}\nabla\times\nabla\times\mathbf{u}=\tilde{\rho}\frac{\partial^2\mathbf{u}}{\partial t^2}. \tag{22}$$

We represent the displacement vector $\mathbf{u}$ in terms of the sum of the potentials, as follows:

$$\mathbf{u}=\nabla\Phi+\boldsymbol{\nabla}\times\boldsymbol{\Psi}, \tag{23}$$

for which the vector potential $\boldsymbol{\Psi}$ satisfies $\boldsymbol{\nabla}\cdot\boldsymbol{\Psi}=0$.

Up to second-order approximation, the scalar potential Φ may be written as

$$\Phi\approx\Phi^{(1)}+\Phi^{(2)}+\text{c.c.}, \tag{24}$$

where

$$\Phi^{(1)}(\mathbf{r},t)=\Phi_1(\mathbf{r})\exp(i\omega t), \tag{25a}$$

$$\Phi^{(2)}(\mathbf{r},t)=\Phi_1(\mathbf{r})\exp(i2\omega t). \tag{25b}$$

Substitution of Eqs. (23) and (24) in Eq. (22) yields

$$\tilde{C}_{11}^{(1)}\nabla^2\Phi^{(1)}=\tilde{\rho}\frac{\partial^2\Phi^{(1)}}{\partial t^2}, \tag{26a}$$

$$\tilde{C}_{11}^{(2)}\nabla^2\Phi^{(2)}=\tilde{\rho}\frac{\partial^2\Phi^{(2)}}{\partial t^2}, \tag{26b}$$

$$\tilde{\mu}\nabla^2\boldsymbol{\Psi}=\tilde{\rho}\frac{\partial^2\boldsymbol{\Psi}}{\partial t^2}. \tag{27}$$

As observed from Eq. (27), this equation takes on a nonresonant form and the influence of the existence of the bubbles on the propagation of the shear wave in a bubbly soft medium is insignificant. In the following we shall restrain our attention in the propagation of the compressional wave in such a medium.

Substituting Eqs. (20), (21), and (25) in Eq. (26), one arrives at the equations that must be satisfied by the scalar potentials of the first and the second order, as follows:

$$C_{11}\nabla^2\Phi_1+\tilde{\rho}\omega^2(1+V_{g1}C_{11})\Phi_1=0, \tag{28a}$$

$$C_{11}\nabla^2\Phi_2+4\tilde{\rho}\omega^2(1+V_{g2}C_{11})\Phi_2=C_{11}V_\Gamma(\tilde{C}_{11}^{(1)})^2(\nabla^2\Phi_1)^2. \tag{28b}$$

Equations (28a) and (28b) give description of the propagation of the fundamental and the second harmonics of the compressional wave in a bubbly soft medium, respectively. Note that Eq. (28) is derived on the basis of Eq. (22), which is expressed as a form of a linear order terms with nonlinear propagation parameters due to the nonlinear oscillation of bubbles. Consequently it is seen that Eq. (28b) takes a simple form without any quadratic term involved that represents the dynamic nonlinearity caused by the finite amplitude of wave. On the right-hand side of this equation, however, a quadratic term appears that accounts for the transfer of acoustical energy from the fundamental to the second harmonic waves, which results from the strong physical nonlinearity that dominates for a bubbly medium.

D. One-dimensional case

Now consider a one-dimensional case in which a plane longitudinal wave propagates along the $\hat{x}_1$ direction in a bubbly soft medium. For simplicity, without losing generality, all the bubbles are assumed to have the same equilibrium radius R_0. In such a case Eq. (25) becomes

$$\Phi^{(1)}(x_1,t)=\Phi_1(x_1)\exp(i\omega t), \tag{29a}$$

$$\Phi^{(2)}(x_1,t)=\Phi_2(x_1)\exp(i2\omega t). \tag{29b}$$

Using the Kelvin–Voigt viscoelastic model, the Lamé coefficients of the soft viscoelastic medium may be rewritten as

$$\lambda=\lambda_m,\quad \mu=\mu_m+\eta_m\partial/\partial t. \tag{30}$$

Substitution of Eqs. (29) and (30) in Eq. (28) yields

$$\frac{d^2\Phi_1}{dx_1^2}+(\Lambda_1^2+\tilde{\rho}\omega^2V_{g1})\Phi_1=0, \tag{31a}$$

$$\frac{d^2\Phi_2}{dx_1^2}+(\Lambda_2^2+4\tilde{\rho}\omega^2V_{g2})\Phi_2=\tilde{\rho}^2\omega^4V_\Gamma\Phi_1^2, \tag{31b}$$

where $\Lambda_1^2=\omega^2\tilde{\rho}/[(\lambda_m+2\mu_m)+i2\omega\eta_m]$ and $\Lambda_2^2=4\omega^2\tilde{\rho}/[(\lambda_m+2\mu_m)+i4\omega\eta_m]$.

The effective wave numbers are defined as complex numbers that can be expressed in terms of real effective wave speeds and effective attenuations, as follows:

$$\tilde{k}_1=\sqrt{\Lambda_1^2+\tilde{\rho}\omega^2V_{g1}}=\omega/c_1-i\alpha_1,$$

$$\tilde{k}_2=\sqrt{\Lambda_2^2+4\tilde{\rho}\omega^2V_{g2}}=2\omega/c_2-i\alpha_2,$$

where $\tilde{k}_i$ refers to the effective wave numbers, c_i and α_i refer to the (real) effective wave speed and the effective attenuation, respectively; and the subscripts $i=1$ and 2 refer to the fundamental wave and the second harmonic wave, respectively.

Assuming $\Phi_1=\Phi_{1A}\exp(-i\tilde{k}_1x_1)$, from Eq. (31a) one readily obtains the expressions of c_1 and α_1, as follows:

$$c_1=\left(\frac{-A_1+\sqrt{A_1^2+B_1^2}}{2}\right)^{-1/2}, \tag{32a}$$

$$\alpha_1=\omega B_1c_1/2, \tag{32b}$$

where the parameters of A_1 and B_1 are given in the Appendix.

It is apparent that the solution of Eq. (31b) is supposed to consist of a general solution and a special solution. By invoking the boundary condition that the second harmonic wave should be zero at the beginning, i.e., $\Phi_2|_{x_1=0}=0$, one can readily determine the expression of the second harmonic wave as follows:

$$\Phi_{2g} = C\Phi_{1A}^2[\exp(-i\tilde{k}_2 x) - \exp(-i\tilde{k}_2 x)] + \text{c.c.}, \tag{33}$$

where

$$c_2 = \left(\frac{-A_2 + \sqrt{A_2^2 + B_2^2}}{2}\right)^{-1/2}, \tag{34a}$$

$$\alpha_2 = \omega B_2 c_2, \tag{34b}$$

where the parameters of A_2, B_2, and C are given in the Appendix.

In practice, the nonlinearity parameter (B/A) is of particular significance that may be used to define the nonlinearity of the media. In the present study, therefore, an effective nonlinearity parameter $(B/A)_e$ is introduced to describe the extent to which the nonlinearity of a bubbly medium is enhanced by the nonlinear oscillation of bubbles. The value of $(B/A)_e$ may be determined near the natural frequency of the bubble, as given in[14]

$$(B/A)_e = \left|\frac{4C\rho_m c_m^3(2\alpha_1 - \alpha_2)}{\omega} - 2\right|.$$

It is apparent that the expression of the effective nonlinearity parameter $(B/A)_e$, derived here, is identical in form with the one obtained by Ma *et al.*,[14] except that their approach only applies to a liquid containing shelled bubbles.

IV. NUMERICAL RESULTS

In this section, we shall exemplify numerically the application of the EMM, as well as some previous methods, in three particular cases that are chosen to represent the important practical situations and are verifiable by actual experiments. A series of comparisons have been performed between the numerical results given by the EMM and by the other ones, for the purpose of inspecting the efficiency and accuracy of the EMM in studying the sound propagation in a bubbly soft medium. The typical results of the effective acoustical parameters of such media are demonstrated and analyzed in detail to come to the conclusion that the EMM is valid and can give more precise results when the influences of the additional effects are not negligible.

A. Blood with contrast agents

Consider the sound propagation in the blood with the injection of contrast agents. In this situation, one has $\mu_m = 0$ and the magnitude of compressibility term is negligible for small amplitude pulsation of bubbles. Then, it is seen that Eq. (14) is equivalent to the Rayleigh–Plesset-like equation derived by Church [see Eq. (17) in Ref. 20 where the variable is the radius variation], and the effective acoustical parameters given by Eqs. (32) and (34) are identical in form with those given by the method of Ma *et al.*,[14] except for the fact that the effect of ambient pressure has been included in the present model.

For practical clinic application of nonlinear ultrasound imaging utilizing the contrast agents, it is crucial to accurately predict the effective acoustical parameters for guaranteeing the precision of the ultrasound imaging of the blood flow. In the presence of contrast agents, the effective parameters of blood will inevitably be affected by the change in blood pressures that serve as the ambient pressures. Indeed, the actual values of human blood pressures may vary within a considerably large range when compared with the atmospheric pressure. For a healthy human the normal values of the systolic blood pressure (SBP) and the diastolic blood pressure (DBP) are approximately 120 and 70 mm Hg, respectively. However, the value of SBP may generally reach 150 mm Hg (even exceed 200 mm Hg in some severe cases) for hypertensive patients, and the value of DBP may be less than 50 mm Hg for hypotensive patients. Therefore, it is significant to incorporate the effect of the ambient pressure properly for obtaining a precise result, which could not be accomplished by only using the approach of Ma *et al.*[14]

We calculate the effective acoustical parameters for different values of blood pressures and present the results in Fig. 2. In the numerical experiments, three particular values of the blood pressure P_b are used: $P_b = 76$ mm Hg (solid line), 140 mm Hg (dashed line), and 50 mm Hg (dotted line). For comparison the method of Ma *et al.*[14] is also employed in the case of $P_b = 76$ mm Hg (i.e., $P_b = P_g = 1$ atm) and the results are represented by the circles in Fig. 2. The contrast agent is chosen as Albunex® diluted so that $\beta = 0.003\%$, for which the parameters are $R_0 = 1.7$ μm, $h = 20$ nm, $\mu_m = 55.6$ MPa, $\eta_m = 2.2$ Pa s, $\sigma_s = 0.04$ N/m.[14] For blood the parameters are assumed to be $\rho_m = 1060$ kg/m^3, $c_m = 1540$ m/s, and $\sigma_m = 0.056$ N/m.[21] Observation of Fig. 2 proves that the results predicted by using the EMM agree well with the method of Ma *et al.*[14] in the case where $P_g = P_b$. In the cases where $P_g \neq P_b$, it is obvious that the effective parameters are notably affected by the blood pressures. As the ambient pressure increases, the bubble resonance frequencies tend to increase due to the reduction of the equilibrium radii of bubbles, and it is more difficult for the bubbles to pulsate, which results in a decrease of the change in the effective wave speed as well as the magnitude of the attenuation peaks. In such cases, neglect of the contributions of the ambient pressure definitely results in nontrivial errors, and it is fair to conclude that the EMM gives more precise results than in the method of Ma *et al.*[14] However, we must stress that the validity of such a statement is certainly not absolute. Indeed, the EMM as well as the method of Ma *et al.*[14] are developed on the basis of the theoretical models that only account for those principal physical mechanisms, but such simplification of the model will inevitably lose some particular mechanisms that could be significant in practical situations. For a typical example, in the present study and the study of and Ma *et al.*[14] bubbles are assumed to keep spherical shapes approximately under the action of acoustic pressures, whereas the possible buckling effects of the bubbles have not been considered that may notably affect the results under certain circumstances.[17] When the effect of buckling of bubbles dominates, which occurs as the hydrostatic pressure reaches a critical value, it has been proved through experiments that the value of the effective nonlinearity parameter of liquid containing contrast agents will be very large and negative.[17] In such cases it is apparently not reasonable to apply unconditionally either the EMM or the method of

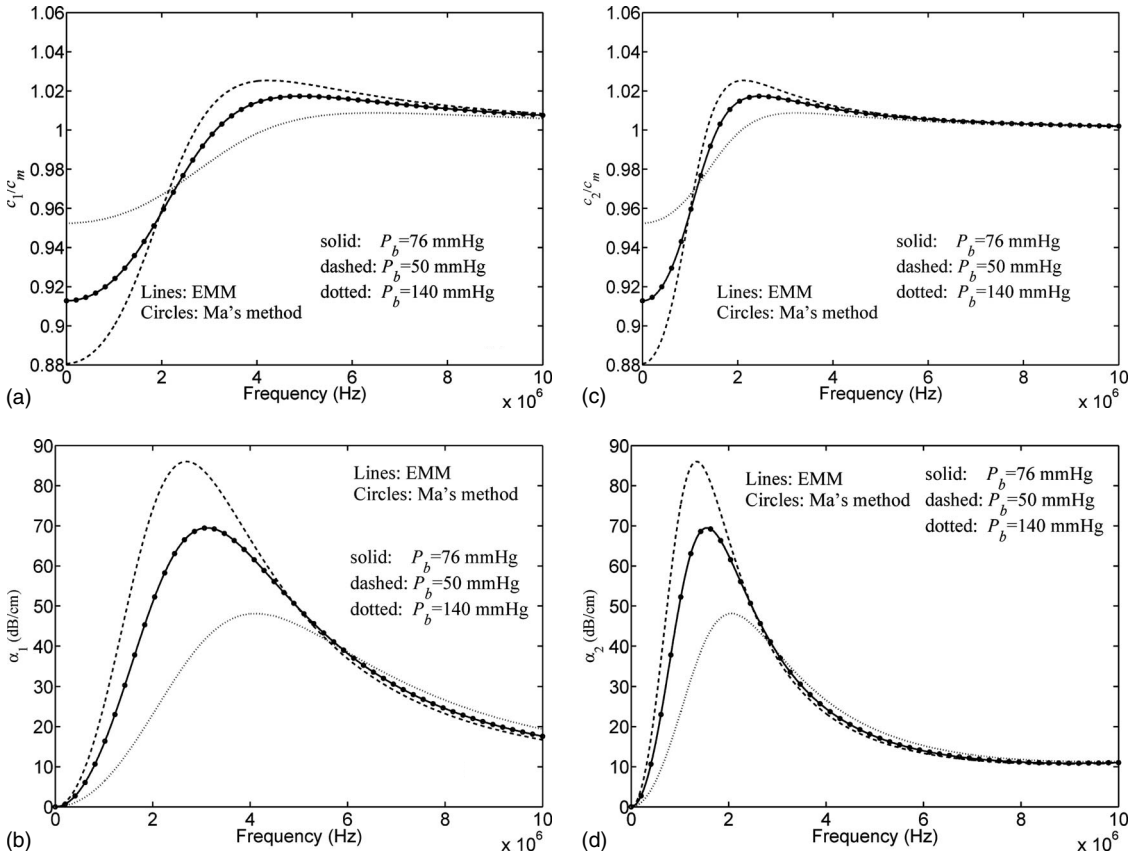


FIG. 2. Comparison of the dispersive [(a) and (c)] and the attenuation [(b) and (d)] curves in blood with contrast agents. (a) and (b) Fundamental wave and (c) and (d) second harmonic wave.

Ma *et al.*[14] at low frequencies. It is important to keep the limitations of the EMM and the previous methods in mind as one goes to the following study.

B. A soft medium containing contrast agents

Consider the propagation of a longitudinal wave in a soft medium containing contrast agents that may be regarded as a simple model of the important application of nonlinear acoustics, for instance a soft tissue in which the contrast agents are introduced to significantly enhance the nonlinearity of the resulting composite medium. In such cases the bubbly soft medium is generally subject to initially prestressed states that may result from the hydraulic pressure or blood pressure, etc. Then it may be readily proved that the present model takes an identical form with the one developed by Qin *et al.*,[16] except that their method could not take into account the existence of the encapsulating shell of contrast agents.

A series of numerical experiments have been carried out and the typical results are presented in Fig. 3. In Fig. 3 the soft medium is chosen as gelatin for which the mechanical parameters are $\rho_m=1000\ \mathrm{kg/m^3}$, $\lambda_m=2.3$ GPa, $\mu_m=0.1$ MPa, $\eta_m=1$ Pa s, and $\sigma_m=0.05$ N/m,[22] and the contrast agents is Albunex® which is diluted so that $\beta=0.005\%$. Another case is also considered to exemplify the effect of shell, where the bubbles are assumed free bubbles of the same radii as Albunex® (i.e., $h=0$). For the two cases the bubbly soft media are subject to a prestressed state with $P_\infty=2$ atm. It is apparently seen that the two methods are exactly interchangeable for the case where $h=0$, whereas the results given by the two methods notably differ from each other for a soft medium containing contrast agents. In the presence of shells, the bubbles tend to appear acoustically smaller than they actually are, and the pulsation amplitude of the bubble oscillation is restrained by the shell.[20] This leads to the fact that for the contrast agents, the resonance frequency tends to increase, and the magnitude of the dispersive and the attenuation curves are suppressed which result from the resonance mechanism of bubbles.

C. Soft medium permeated with free bubbles

Consider the propagation of a longitudinal wave in a soft medium permeated with free bubbles that applies to a variety of important situations. One typical application is the resonance sound absorbent. Assuming an initially unstrained case, one has

$$\frac{\mu_m}{\mu_s}=\frac{\rho_m}{\rho_s}=\frac{\eta_m}{\eta_s}=\frac{c_m}{c_s}=\frac{R}{R_s}=\frac{R_1}{R_{s1}}=\frac{R_0}{R_{s0}}=1.$$

For such a case, it may be easily proved that Eq. (14) and the effective acoustical parameters given by Eqs. (32) and (34) are identical in form, respectively, with the Rayleigh–Plesset-like equation developed by Ostrovsky[3] and the results given by the method of Liang and Cheng,[15] except that the effect of surface tension has been included in the EMM and a difference occurs in the viscous terms. Note that

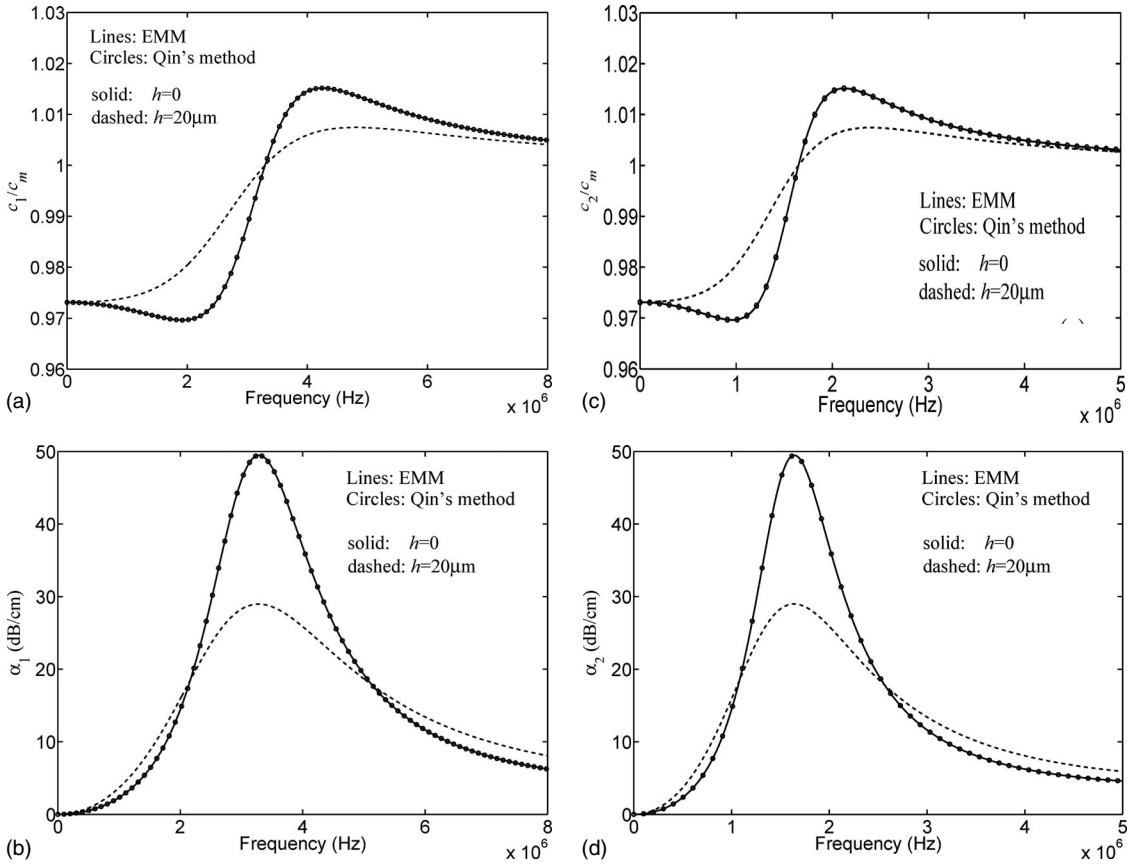


FIG. 3. Comparison of the dispersive [(a) and (c)] and the attenuation [(b) and (d)] curves in gelatin containing contrast agents. (a) and (b) Fundamental wave and (c) and (d) second harmonic wave.

the term $\delta_2 UdU/dt$ accounting for the viscous damping at quadratic order, has been incorrectly missed in the equation of Ostrovsky[3] and thereby, in the method of Liang and Cheng,[15] which was developed on the basis of the model of Ostrovsky.[3] An elastic plastizole[3] is thus employed as the host medium for which the mechanical parameters are $\rho_m = 1000\ \mathrm{kg/m^3}$, $\lambda_m = 2.3$ GPa, $\mu_m = 11.5$ kPa, and $\sigma_m = 0.5\ \mathrm{N/m}$. The bubble parameters are chosen as $R_0 = 50\ \mu$m, $\beta = 0.0005\%$ so that the typical work frequency of the sound absorbent is approximately 70 kHz. The results of the numerical calculations are presented in Fig. 4. It is obvious that for such a case the surface tension acts as the principal effect that affects the accuracy of the results due to the small radii of bubbles. Therefore, we have also studied a case in which the surface tension is ignored by using the EMM as well as the method of Liang and Cheng[15] for a comparison. It is shown by the results that the two methods can give identical results in the case where $\sigma_m = 0$ and in the presence of surface tension, the resonance frequency of the bubbles is reduced as suggested by Zabolotskaya *et al.*[6] Moreover, the magnitude of the peaks in the dispersive and the attenuation curves also increase due to the fact that the surface tension enhances both the linear pulsation and the nonlinear response of bubbles. In such a case the effect of surface tension must be considered to properly evaluate the efficiency of the sound absorbent, and the EMM gives a more precise result than the method of Liang and Cheng.[15]

V. CONCLUSION

We present an EMM for investigating the propagation of the elastic wave in a soft viscoelastic medium permeated with encapsulated bubbles. The EMM enables the investigation of the strong nonlinearity of such a medium by giving description of the second harmonic wave which stems from the nonlinear oscillation of bubbles, and accounts for the effects of weak compressibility, viscosity, surrounding pressure, surface tension, and encapsulating shells. Based on the modified equation of bubble oscillation, the linear and the nonlinear wave equations are derived and solved for a simplified one-dimensional case. Via comparisons of the numerical results given by the EMM and the previous theories in three representative examples, the EMM is exactly valid and can give more precise evaluations of the effective acoustical parameters of the bubbly media when the influences of the additional effects are not negligible. However, it is noteworthy that there also exist limitations in the present model which prevent the EMM from giving accurate results for a bubbly soft medium. This will be the focus of our further investigation.

ACKNOWLEDGMENTS

The reviewers are gratefully appreciated for valuable comments and suggestions that indeed improved the quality of the manuscript. The work is supported by Ministry of Education of China under Grant Nos. 705017 and

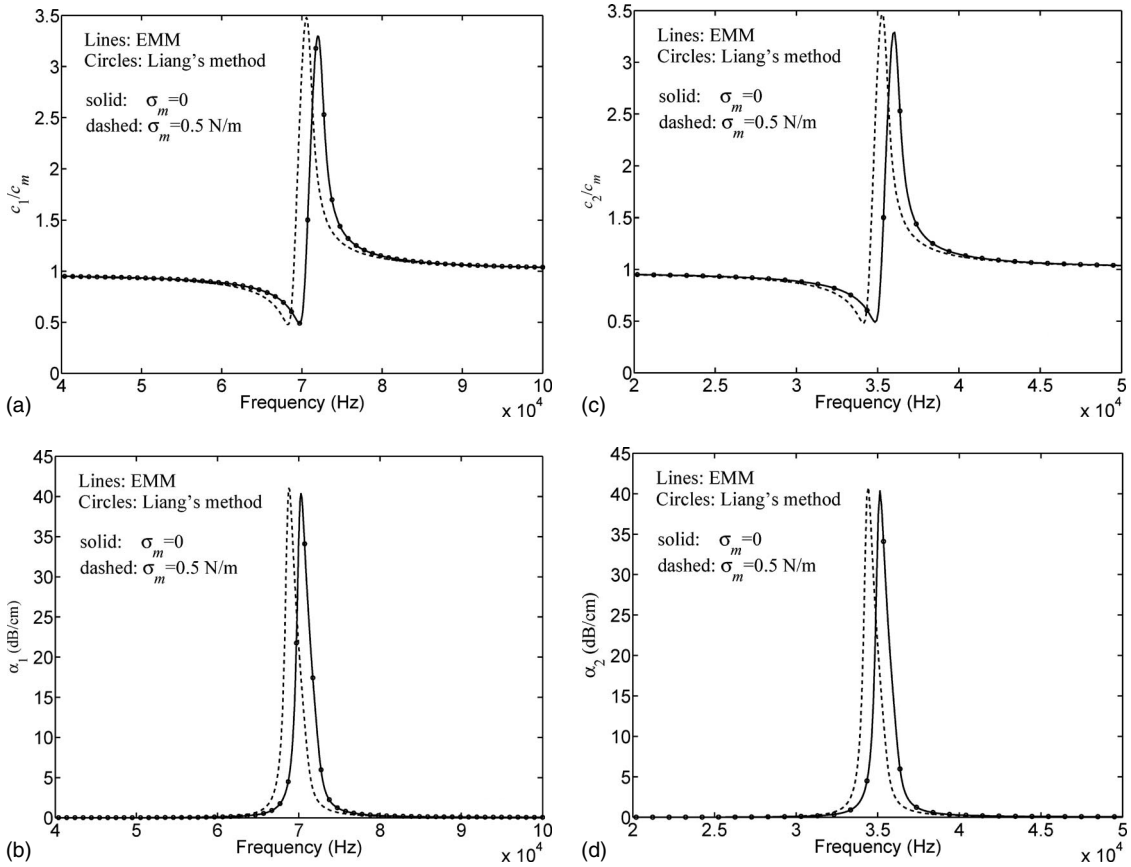


FIG. 4. Comparison of the dispersive [(a) and (c)] and the attenuation [(b) and (d)] curves in elastic plastizole permeated with free bubbles. (a) and (b) Fundamental wave and (c) and (d) second harmonic wave.

20060284035, the Excellent Youth Science Foundation of China under Grant No. 10125417, and the State Key Development Program of Basic Research of China under Grant No. 51315.

APPENDIX

The parameters of A_1 and B_1 used in Eq. (32) are given as follows:

$$A_1 = \tilde{\rho}\omega^2\left[\frac{\lambda_m + 2\mu_m}{(\lambda_m + 2\mu_m)^2 + 4\omega^2\eta_m^2} - \frac{Ne_1(\omega_1^2 - \omega^2)}{(\omega_1^2 - \omega^2)^2 + (\omega\delta_1 + \omega^3 R_1/c_m)^2}\right],$$

$$B_1 = \tilde{\rho}\omega^2\left[\frac{2\omega\eta_m}{(\lambda_m + 2\mu_m)^2 + 4\omega^2\eta_m^2} - \frac{Ne_1(\omega\delta_1 + \omega^3 R_1/c_m)}{(\omega_1^2 - \omega^2)^2 + (\omega\delta_1 + \omega^3 R_1/c_m)^2}\right].$$

The parameters of A_2 and B_2 used in Eq. (34) are given as follows:

$$A_2 = \tilde{\rho}\omega^2\left[\frac{\lambda_m + 2\mu_m}{(\lambda_m + 2\mu_m)^2 + 16\omega^2\eta_m^2} - \frac{Ne_1(\omega_1^2 - 4\omega^2)}{(\omega_1^2 - 4\omega^2)^2 + (2\omega\delta_1 + 8\omega^3 R_1/c_m)^2}\right],$$

$$B_2 = \tilde{\rho}\omega^2\left[\frac{4\omega\eta_m}{(\lambda_m + 2\mu_m)^2 + 16\omega^2\eta_m^2} - \frac{Ne_1(2\omega\delta_1 + 8\omega^3 R_1/c_m)}{(\omega_1^2 - 4\omega^2)^2 + (2\omega\delta_1 + 8\omega^3 R_1/c_m)^2}\right].$$

The parameters of C used in Eq. (34) are given as follows:

$$C = [(M_1N_1 + M_2N_2) + i(M_2N_1 - M_1N_2)]/(N_1^2 + N_2^2),$$

where

$$M_1 = \tilde{\rho}^2\omega^4 Ne_1^2[(G_1 - 3\omega^2 H_1)(A_1^2 - B_1^2) + 2A_1B_1\omega\delta_2],$$

$$M_2 = \tilde{\rho}^2\omega^4 Ne_1^2[(A_1^2 - B_1^2)\omega\delta_2 - 2A_1B_1(G_1 - 3\omega^2 H_1)],$$

$$N_1 = K_1(-4A_1 + A_2) + L_1(B_2 - 4B_1),$$

$$N_2 = L_1(-4A_1 + A_2) - K_1(B_2 - 4B_1),$$

where

$$K_1 = (\omega_1^2 - 4\omega^2)[(\omega_1^2 - \omega^2)^2 - (\omega\delta_1 + \omega^3 R_1/c_m)^2] - 2(\omega_1^2 - \omega^2)(\omega\delta_1 + \omega^3 R_1/c_m)(2\omega\delta_1 + 8\omega^3 R_1/c_m),$$

$$L_1 = [(\omega_1^2 - \omega^2)^2 - (\omega\delta_1 + \omega^3 R_1/c_m)^2](2\omega\delta_1 + 8\omega^3 R_1/c_m) + 2(\omega_1^2 - 4\omega^2)(\omega_1^2 - \omega^2)(\omega\delta_1 + \omega^3 R_1/c_m).$$

[1]E. Meyer, K. Brendel, and K. Tamm, "Theory of resonant scattering from sphereical bubbles in elastic and viscoelastic media," J. Acoust. Soc. Am. **30**, 1116–1124 (1958).
[2]A. C. Eringen and E. S. Suhubi, *Elastodynamics* (Academic, New York, 1974).
[3]L. A. Ostrovsky, "Nonlinearity acoustics of slightly compressible porous media," Sov. Phys. Acoust. **34**, 523–526 (1988).
[4]L. A. Ostrovsky, "Wave processes in media with strong acoustic nonlinearity," J. Acoust. Soc. Am. **90**, 3332–3337 (1991).
[5]S. Y. Emelianov, M. F. Hamilton, Yu. A. Ilinskii, and E. A. Zabolotskaya, "Nonlinear dynamics of a gas bubble in an incompressible elastic medium," J. Acoust. Soc. Am. **115**, 581–588 (2004).
[6]E. A. Zabolotskaya, Yu. A. Ilinskii, G. D. Meegan, and M. F. Hamilton, "Modification of the equation for gas bubble dynamics in a soft elastic medium," J. Acoust. Soc. Am. **118**, 2173–2181 (2005).
[7]B. Liang and J. C. Cheng, "Acoustic localization in weakly compressible elastic medium containing air bubbles," Phys. Rev. E **75**, 016605 (2007).
[8]P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).
[9]G. C. Gaunaurd and H. Überall, "Resonance theory of the effective properties of perforated solids," J. Acoust. Soc. Am. **71**, 282–295 (1982).
[10]G. C. Gaunaurd and J. Barlow, "Matrix viscosity and bubble-size distribution effects on the dynamic effective properties of perforated elastomers," J. Acoust. Soc. Am. **75**, 23–34 (1984).
[11]G. C. Gaunaurd and W. Wertman, "Comparison of effective medium theories for inhomogeneous continua," J. Acoust. Soc. Am. **85**, 541–554 (1989).
[12]A. M. Baird, F. H. Kerr, and D. J. Townend, "Wave propagation in a viscoelastic medium containing fluid-filled microspheres," J. Acoust. Soc. Am. **105**, 1527–1538 (1999).
[13]D. G. Aggelis, S. V. Tsinopoulos, and D. Polyzos, "An iterative effective medium approximation (IEMA) for wave dispersion and attenuation predictions in particulate composites, suspensions and emulsions," J. Acoust. Soc. Am. **116**, 3343–3452 (2004).
[14]J. Ma, J. F. Yu, Z. M. Zhu, X. F. Gong, and G. H. Du, "Acoustic nonlinearity of liquid containing encapsulated microbubbles," J. Acoust. Soc. Am. **116**, 186–193 (2004).
[15]B. Liang and J. C. Cheng, "Propagation of acoustic wave in viscoelastic medium permeated with air bubbles," Chin. Phys. **15**, 412–421 (2006).
[16]B. Qin, J. J. Chen, and J. C. Cheng, "Influence of the surrounding pressure on acoustic properties of slightly compressible media permeated with air-filled bubbles," Acoust. Phys. **52**, 490–496 (2006).
[17]D. H. Trivett, H. Pincon, and P. H. Rogers, "Investigation of a three-phase medium with a negative parameter of Nonlinearity," J. Acoust. Soc. Am. **119**, 3610–3617 (2006).
[18]M. Mooney, "A theory for large elastic deformation," J. Appl. Phys. **11**, 582–597 (1940).
[19]L. D. Landau and E. M. Lifshits, *Theory of Elasticity* (Pergamon, Oxford, 1986).
[20]C. C. Church, "The effects of an elastic solid surface layer on the radial pulsation of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[21]R. E. Apfel and C. K. Holland, "Gauging the likelihood of cavitation from short-pulse, low-duty cycle diagnostic ultrasound," Ultrasound Med. Biol. **17**, 179–185 (1991).
[22]S. Catheline, J. L. Gennisson, and M. Fink, "Measurement of elastic nonlinearity of soft solid with transient elastography," J. Acoust. Soc. Am. **114**, 3087–3091 (2003).

Finite-difference time-domain synthesis of infrasound propagation through an absorbing atmosphere

C. de Groot–Hedlin
Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093

(Received 5 June 2007; revised 16 June 2008; accepted 18 June 2008)

Equations applicable to finite-difference time-domain (FDTD) computation of infrasound propagation through an absorbing atmosphere are derived and examined in this paper. It is shown that over altitudes up to 160 km, and at frequencies relevant to global infrasound propagation, i.e., 0.02–5 Hz, the acoustic absorption in dB/m varies approximately as the square of the propagation frequency plus a small constant term. A second-order differential equation is presented for an atmosphere modeled as a compressible Newtonian fluid with low shear viscosity, acted on by a small external damping force. It is shown that the solution to this equation represents pressure fluctuations with the attenuation indicated above. Increased dispersion is predicted at altitudes over 100 km at infrasound frequencies. The governing propagation equation is separated into two partial differential equations that are first order in time for FDTD implementation. A numerical analysis of errors inherent to this FDTD method shows that the attenuation term imposes additional stability constraints on the FDTD algorithm. Comparison of FDTD results for models with and without attenuation shows that the predicted transmission losses for the attenuating media agree with those computed from synthesized waveforms. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2959736]

PACS number(s): 43.28.Dm, 43.28.Js, 43.20.Hq [VEO] Pages: 1430–1441

I. INTRODUCTION

Ever since pressure fluctuations generated by the 1883 explosion of Krakatoa were recorded on barometers throughout the US, Europe, and Russia (Bass *et al.*, 2006), it has been understood that volcanoes excite infrasonic waves—subaudible acoustic energy—detectable at ranges from hundreds to thousands of kilometers. Over the years, pressure waves generated by both earthquakes and volcanoes have been detected at distances of hundreds to thousands of kilometers from the source regions (e.g., Mikumo, 1968; Wilson and Forbes, 1969; Young and Greene, 1982). Currently, a global infrasound network consisting of 60 infrasound stations distributed over the Earth is being deployed as one part of the International Monitoring System (IMS) designed to detect nuclear explosions (Der *et al.*, 2002). The construction of this network of stations, which measure differential pressure in the frequency range from roughly 0.02 to 5 Hz (e.g., Arrowsmith and Hedlin, 2005), has allowed infrasound to emerge as a method of investigating a much broader range of atmospheric and geophysical phenomena than was possible until now. Recent work has focused on such disparate infrasound sources as meteors (ReVelle *et al.*, 2004), ocean swells (Garceś *et al.*, 2004a), surf noise (Arrowsmith and Hedlin, 2005), and tsunamis (Le Pichon *et al.*, 2005), as well as the tragic Columbia shuttle explosion (Garceś *et al.*, 2004b). In other studies, seasonal variations in infrasound arrivals from recurring sources were used to probe atmospheric properties (e.g., Le Pichon *et al.*, 2002; Garceś *et al.*, 2004a). However, improvements are needed in both atmospheric models and acoustic propagation modeling methods in order to correctly interpret the wide range of signals detected by the global infrasound network (Bass *et al.*, 2006).

Infrasound propagation is controlled mainly by reflections by the ground terrain, scattering by atmospheric turbulence, advection through a windy atmosphere, refraction due to sound-speed gradients, and attenuation due to the absorbing properties of the atmosphere. Infrasound detections at long ranges from the source are generally either thermospheric arrivals, which result from downward refraction by the steep temperature gradients of the upper atmosphere, or stratospheric arrivals, which are ducted between the ground and the stratopause. Detection of the latter depends strongly on the stratospheric winds, while thermospheric waves are strongly attenuated due to enhanced atmospheric absorption within the thermosphere. Finite-difference time-domain (FDTD) methods have been developed to model the effects of sound propagation in a windy atmosphere (e.g., Ostashev *et al.*, 2005). The goal of the present study is to develop a FDTD method of simulating infrasound propagation through a heterogeneous, absorptive atmosphere.

The effects of absorption are usually simulated using frequency-domain modeling techniques, such as parabolic equation (PE) or fast field program (FFP) methods (e.g., Jensen *et al.*, 1994). These methods use a heuristic approach of adding a small imaginary value to the acoustic velocity to simulate attenuation. However, these frequency-domain methods make use of approximations to the atmospheric model and/or to the governing wave equation that limit their applicability. For instance, FFP methods assume that material properties of the model vary only in a single spatial coordinate, while PE methods compute one-way propagation, which is valid only for models with limited range dependence.

By contrast, the FDTD approach is applicable to a wide range of complicated phenomena in heterogeneous media as

it relies on a discretization of the governing equations. The wider applicability of the FDTD method comes at the cost of greater computational time and memory requirements. However, with advances in computing technology, the FDTD method has drawn greater interest. A method of incorporating attenuation into an FDTD approach was developed by Day and Minster (1984) for anelastic media, with refinements by Emmerich and Korn (1987), Blanch *et al.* (1995), and Day (1998). These methods approximate the frequency-domain stress-strain relations by low-order rational functions of frequency. The approach may be applied to many types of anelastic media. However, a drawback is that this approach is computationally time- and memory-intensive as compared to nonabsorptive materials.

In the next section, sound absorption mechanisms within the atmosphere are described. It is shown that at frequencies relevant to global infrasound investigations, i.e., 0.02–5 Hz, the attenuation coefficient (in dB/m) can be decomposed into two parts: a constant term and another proportional to the square of the frequency. In Sec. III, it is demonstrated that these attenuation effects may be simulated by modifying the wave equation to include two attenuation terms. The degree to which the revised wave equation introduces material dispersion is examined over the infrasound frequency bandwidth for realistic attenuation values. In Sec. IV, two sets of partial differential equations that are first order in time are introduced to replace the attenuating wave equation. A numerical algorithm to implement these FDTD equations is developed for the case of two-dimensional (2D) models. It is shown that inclusion of an attenuation term proportional to the square of the frequency imposes additional constraints on the model discretization in order to maintain numerical stability. In Sec. V, the algorithm is applied to several environmental models, and FDTD solutions computed with and without atmospheric absorption are compared. The paper concludes with a discussion and summary of results.

II. ATMOSPHERIC ABSORPTION OF SOUND

The absorption of sound within any gas is controlled primarily by the composition of the gas mixture, the ambient pressure, temperature, sound frequency and, at higher frequencies, by the humidity (Bass *et al.*, 1995; Dain and Lueptow, 2001; Sutherland and Bass, 2004). Absorption mechanisms can be categorized in terms of classical losses caused by the irreversible transfer of the kinetic energy of the acoustic wave into heat, and molecular relaxation losses, which involve the excitation of the energy states of the gas molecules comprising the atmosphere (Evans *et al.*, 1972; Ejakov *et al.*, 2003). Classical absorption includes losses due to diffusion, heat conduction, and viscosity; molecular relaxation losses must be computed for each of the primary gas constituents in the air–O_2, N_2, CO_2, and O_3. The absorption mechanisms are cumulative, thus, the total attenuation coefficient of sound, in dB/m, is given by the linear sum of the attenuation components associated with each individual loss mechanism.

Attenuation coefficients computed using the expressions given in Sutherland and Bass (2004, 2006) are shown as a function of altitude for frequencies ranging from 0.05–4 Hz in Fig. 1(a). A simplified form of the attenuation coefficient is sought as a function of frequency at each altitude for the purpose of accurately including attenuation effects within the finite-difference computations. For infrasonic frequencies, classical losses dominate at altitudes above approximately 60 km, and the associated attenuation coefficient, in dB/m, is proportional to the square of the frequency (Sutherland and Bass, 2004). At lower altitudes, losses associated with excitation of vibrational modes of the gas molecules dominate. For each of the gas constituents in the air, these losses, $q_{\mathrm{vib},i}$ are proportional to

$$\frac{(f^2/f_{\mathrm{vib},i})}{[1+(f/f_{\mathrm{vib},i})^2]c} \tag{1}$$

(Sutherland and Bass, 2004), where f is the sound frequency, $f_{\mathrm{vib},i}$ is the relaxation frequency for a particular gas component, and c is the sound speed in the absence of attenuation. The attenuation coefficient associated with the excitation of the vibrational modes of each component of the gas, in dB/m, is thus approximately proportional to the square of the frequency at low frequencies $f \ll f_{\mathrm{vib},i}$, and is nearly a constant at high frequencies $f \gg f_{\mathrm{vib},i}$.

The attenuation curves of Fig. 1(a) suggest that the cumulative attenuation coefficient q, expressed in dB/m, may be approximated by decomposing it into a component proportional to the square of the frequency plus a constant at

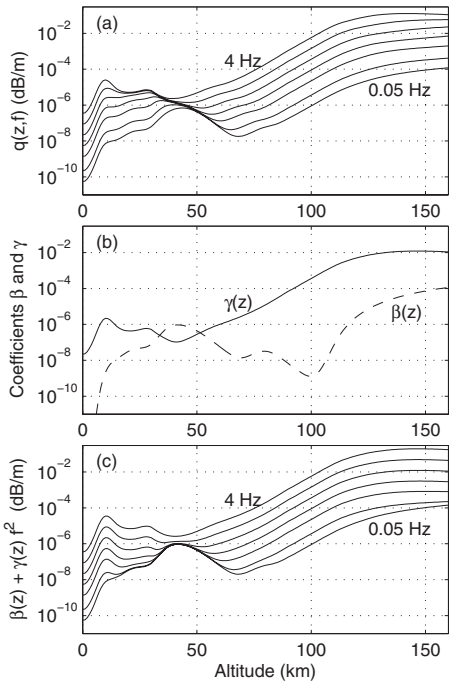


FIG. 1. (a) Attenuation coefficients $q(z,f)$ for frequencies of 0.05 (lowest curve), 0.1, 0.25, 0.5, 1, 2, and 4 Hz (uppermost curve), computed using the expressions given in Sutherland and Bass (2004). (b) Decomposition of the attenuation coefficients at each altitude in terms of a frequency-independent component $\beta(z)$, shown by the dashed line, and a component of the attenuation $\gamma(z)$ proportional to f^2, shown by the solid line; $\beta(z)$ is in units of m^{-1}, γ has units of $\mathrm{s}^2\,\mathrm{m}^{-1}$. Since the decomposition is imprecise, the coefficients depend on the bandwidth chosen; values shown here are derived for the frequency range from 0.05–4 Hz shown in (a). (c) Attenuation curves reconstructed using $\beta(z)+\gamma(z)f^2$.

each altitude, although the fit may be degraded where vibrational losses dominate, as discussed above. Accordingly, the coefficients β and γ that yielded the least-squares best fit to $q(z,f)=\beta(z)+\gamma(z)f^2$ were computed at each altitude. The result of this decomposition for the frequency range 0.05–4 Hz is shown in Fig. 1(b); the solid line shows the derived $\gamma(z)$ attenuation coefficients and the dashed line indicates $\beta(z)$. The attenuation curves in Fig. 1(c) were computed using $\beta(z)+\gamma(z)f^2$; comparison with the curves in Fig. 1(a) indicates the degree of accuracy of this approximation over the frequency bandwidth relevant to infrasound propagation. Discrepancies in the fit are most noticeable at the lower altitudes. Since attenuation is almost negligible within the lower atmosphere—an attenuation coefficient of 10^{-6} dB/m corresponds to a 0.1-dB loss over a 100-km path length—these small errors are insignificant. The misfit between attenuation values shown in Fig. 1(a) and Fig. 1(b) approach 20–30% at altitudes over 120 km. However, attenuation estimates are only approximate at altitudes greater than 90 km due to uncertainties in the atmospheric composition (Sutherland and Bass, 2004). A misfit up to 30% may thus be considered sufficient, given that the attenuation spans many orders of magnitude over the infrasound propagation band.

As indicated in Fig. 1(c), the derived attenuation coefficients $\beta(z)+\gamma(z)f^2$ dB/m are reasonably accurate over approximately two orders of magnitude in frequency. However, these coefficients must be rederived for any other frequency band of interest since the decomposition is inexact. In general, the constant term $\beta(z)$ is larger and the $\gamma(z)$ term is smaller for higher frequency bands.

III. ATMOSPHERIC ABSORPTION AND VELOCITY DISPERSION

Given that the atmospheric absorption of sound may be approximated by an attenuation coefficient with two terms—a constant plus a term proportional to the square of the frequency—an acoustic plane wave propagating in the positive x direction must have pressure fluctuations of the form

$$p(x,t) \propto e^{i2\pi f(x/c_p - t)} e^{-\alpha(f)x}, \tag{2}$$

where c_p is the phase speed, t is time, and the attenuation $\alpha=q/(20\log_{10}(e))$ is a scaled version of the attenuation defined in the previous section.

It will now be shown that the solution to the following equation has the form given by Eq. (2) for the attenuation values and frequencies considered in this paper:

$$c^2\left(1+\mu\frac{\partial}{\partial t}\right)\frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial t^2} + \xi\frac{\partial p}{\partial t}, \tag{3}$$

where μ is the shear viscosity, in units of seconds, and ξ is a damping coefficient, in units of s^{-1}. As outlined in the Appendix (for a 3D medium), this equation describes plane-wave propagation though a homogeneous compressible Newtonian fluid with a small external damping force. The case $\mu\neq 0$, $\xi=0$ is a medium in which internal viscosity is the dominant absorption mechanism (Knopoff, 1956; Lighthill, 1978). For the case $\mu=0$, $\xi\neq 0$, Eq. (3) corresponds to a simple damped oscillator, with a viscous force proportional to the particle velocity opposing the motion (e.g., Boyce and di Prima, 1977). The latter case reduces to the telegraph equation, which combines the wave equation and the diffusion equation. Here, we treat only the case $\xi\partial p/\partial t \ll \partial^2 p/\partial t^2$ so that wave motion dominates, not diffusion.

Inserting a trial solution of the form $p(x,t)=e^{i(kx-\omega t)}$ into Eq. (3), where ω is the circular frequency and k is the wavenumber, yields a complex-valued (squared) wavenumber of the form

$$k^2(\omega) = \frac{1}{c^2}\frac{\omega^2+i\xi\omega}{(1-i\mu\omega)}. \tag{4}$$

From Eq. (2), it can be seen that the wavenumber k is related to the frequency-dependent phase speed $c_p(\omega)$ and attenuation through

$$k = \omega/c_p(\omega) + i\alpha(\omega). \tag{5}$$

By writing $k^2(\omega)$ in exponential notation as $k^2(\omega)=|k^2|\exp(i\phi)$, where ϕ is the argument of $k^2(\omega)$, and applying Eq. (5), it can be seen that

$$\omega/c_p(\omega) = \mathrm{Re(k)} = \sqrt{|k^2|}\cos(\phi/2), \tag{6a}$$

$$\alpha(\omega) = \mathrm{Im}(k) = \sqrt{|k^2|}\sin(\phi/2). \tag{6b}$$

With some algebra manipulations, and applying trigonometric half-angle formulas, it can be shown that

$$\left(\frac{c_p(\omega)}{c}\right)^2 = \frac{2(1+\mu^2\omega^2)}{(1-\mu\xi)+\sqrt{(1+(\xi/\omega)^2)(1+\mu^2\omega^2)}}, \tag{7a}$$

and

$$\alpha^2(\omega) = \frac{\omega^2}{2c^2(1+\mu^2\omega^2)}\left[\sqrt{(1+(\xi/\omega)^2)(1+\mu^2\omega^2)} - (1-\mu\xi)\right]. \tag{7b}$$

Thus, the solution to Eq. (3) is

$$p(x,t) = e^{i\omega(x/c_p(\omega)-t)}e^{-\alpha(\omega)x}, \tag{8}$$

where $c_p(\omega)$ and $\alpha(\omega)$ are the positive square roots of Eqs. (7a) and (7b), respectively, for plane waves propagating in the positive x direction.

For small viscosity coefficients, i.e., $\xi \ll \omega$ and $\mu\omega \ll 1$, and retaining terms up to second order in (ξ/ω) and $\mu\omega$, it can be shown that $\alpha^2(\omega)\approx(\xi+\mu\omega^2)^2/(4c^2)$, so

$$\alpha(\omega) \approx (\xi+\mu\omega^2)/(2c), \tag{9}$$

which has the desired response for acoustic propagation through an absorbing atmosphere. Comparing Eqs. (2) and (9), the coefficient of viscosity μ is related to γ through

$$\mu = \gamma c/(40\pi^2\log_{10}(e)). \tag{10a}$$

and the damping coefficient ξ relates to β as

$$\xi = c\beta/(10\log_{10}(e)). \tag{10b}$$

These values are dependent on altitude, as both c and the attenuation coefficients γ and β vary with altitude. Coeffi-

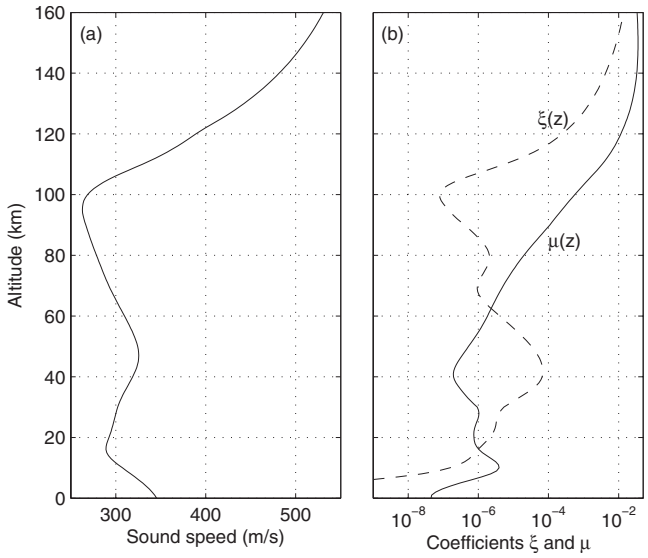


FIG. 2. (a) Sound velocity profile. (b) Coefficients $\xi(z)$ (dashed line, in units of s^{-1}) and $\mu(z)$ (solid line, in units of seconds) corresponding to the given sound velocity profile, and to the coefficients $\beta(z)$ and $\gamma(z)$, shown in Fig. 1(b).

cients $\mu(z)$ and $\xi(z)$ are shown in Fig. 2(b) for the sound velocity profile shown in Fig. 2(a). These coefficients are used in the remainder of this paper to designate realistic atmospheric attenuation value.

Equation (7a) indicates that the acoustic velocity is dispersive; this is in agreement with the Kramers-Kronig relations, initially developed in electromagnetic theory, which show that phase velocity shows frequency dependence in attenuating media. The relation between attenuation and dispersion for acoustic waves in media with attenuation obeying other frequency power laws in discussed in Aki and Richards (1980). More recently, Bass *et al.* (2007) examined the effect of velocity dispersion on sound refraction in the upper atmosphere. Figure 3 shows the derived $c_p(f)/c$ ratios at altitudes of 100, 120, 140, and 160 km for frequencies from 0.01 to 2 Hz, derived using Eq. (7a) and attenuation values shown in Fig. 2(b). These are compared to corresponding

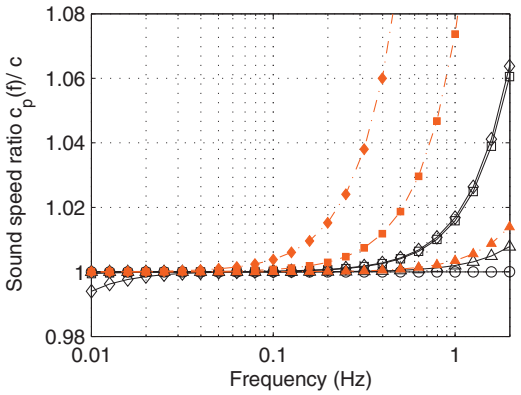


FIG. 3. Ratios of c_p/c as a function of frequency at altitudes of 100 km (circles), 120 km (triangles), 140 km (squares), and 160 km (diamonds). The black lines with open symbols indicate ratios computed using Eq. (7a); the red dash-dot lines with filled symbols indicate corresponding values for 120, 140, and 160 km, computed using a formula from Bass *et al.* (2007). The values of c at these altitudes are shown in Fig. 2(a). At these frequencies, velocity dispersion is significant only at altitudes over 100 km.

ratios computed using formulas found in Bass *et al.* (2007). The results show that dispersion is negligible at altitudes less than 100 km over the range of frequencies typically recorded in global infrasound studies. At higher altitudes, both models indicate significant increases in velocity at the upper end of the infrasound frequency band. However, velocity dispersion at the upper end of the frequency range is more pronounced for the Bass *et al.* (2007) formulas than is predicted by Eq. (7a). It is not clear why this is true. As shown, Eq. (7a) predicts a slight decrease in phase velocity at the lower end of the frequency band at an altitude of 160 km.

IV. ACOUSTIC PROPAGATION IN AN ABSORBING ATMOSPHERE

A. First-order differential equations

As shown in Eqs. (A7), (A85), and (A18) of the Appendix, the following pair of first-order differential equations in time governs linear acoustic propagation through an attenuating atmosphere:

$$\frac{\partial \mathbf{v}_s}{\partial t} = \mathbf{f}_i - \frac{1}{\rho}\left[\nabla p_s - (3\rho c^2/4)\mu\left(\nabla^2 \mathbf{v}_s + \frac{1}{3}\nabla(\nabla \cdot \mathbf{v}_s)\right)\right] - \xi \mathbf{v}_s, \tag{11a}$$

$$\frac{\partial p_s}{\partial t} = -\rho c^2 \nabla \cdot \mathbf{v}_s, \tag{11b}$$

where p_s is defined as the pressure perturbation due to the sound wave, $\mathbf{v}_s$ is the acoustic particle velocity, and ρ is the atmospheric density. The addition of the ξ term is heuristic, added only to yield more realistic atmospheric absorption values. Terms proportional to gravity are important at very low frequencies corresponding to internal waves (Ostashev *et al.*, 2005; Lingevitch *et al.*, 1999) and are omitted here. The effects of wind are beyond the scope of this paper. As expected, Eqs. (11) reduce to the standard FDTD equations for acoustic propagation (Botteldooren, 1994) for the case $\mu=\xi=0$.

As shown in Eq. (A17) in the Appendix, the coupled equations yield

$$\frac{\partial^2 p_s}{\partial t^2} = c^2\left(1 + \mu\frac{\partial}{\partial t}\right)\nabla^2 p_s - \xi\frac{\partial p_s}{\partial t} - \frac{c^2}{\rho}\nabla\rho\cdot\nabla p_s, \tag{12}$$

where the external force has been omitted. For a homogeneous medium, this reduces to

$$\frac{\partial^2 p_s}{\partial t^2} = c^2\left(1 + \mu\frac{\partial}{\partial t}\right)\nabla^2 p_s - \xi\frac{\partial p_s}{\partial t}. \tag{13}$$

In this case the pressure depends only on the distance from the source. The outgoing solution of Eq. (13) for a point source at the origin is given by the real part of

$$p_s(R,t) = A\frac{e^{i\omega(R/c_p(\omega)-t)}}{R}e^{-\alpha(\omega)R}, \tag{14}$$

where $R=\Sigma(x_i x_i)^{1/2}$ is the distance from the source, A is a complex-valued constant, and $c_p(\omega)$ and $\alpha(\omega)$ are the positive square roots of Eqs. (7a) and (7b), respectively.

B. Propagation through a medium with exponentially decreasing density

Equation (14) holds for an isothermal model with constant density and attenuation values. An analytic expression may also be derived for a somewhat more realistic atmospheric model represented as an isothermal medium with exponentially decreasing density. Using the hydrostatic equation $dp_o/dz=-g\rho$, where p_o is the atmospheric pressure in the absence of a sound wave and g is the acceleration due to gravity, and the ideal gas law $p_o=\rho R_a T$, where T is the temperature in degrees Kelvin and $R_a=287.04\ \mathrm{J\ kg^{-1}\ K^{-1}}$ is the gas constant for dry air, the density gradient with altitude is derived as

$$\frac{\partial\rho(z)}{\partial z}=-\frac{\rho g}{R_a T}, \tag{15}$$

for an isothermal atmosphere. Horizontal density gradients are negligible. The atmospheric density decreases by about 6 orders of magnitude over 100 km in altitude.

The density may be defined as $\rho(z)=\rho(0)\ e^{-az}$ for an isothermal atmosphere, where $a=g/R_a T$. Combining this expression for density with Eq. (12) then yields

$$\frac{\partial^2 p_s}{\partial t^2}=c^2\left(1+\mu\frac{\partial}{\partial t}\right)\nabla^2 p_s-\xi\frac{\partial p_s}{\partial t}+ac^2\frac{\partial p_s}{\partial z}. \tag{16}$$

The outgoing solution of Eq. (16) for a point source at the origin is given by the real part of

$$p_s(R,t)=Ae^{-az/2}\frac{e^{i\omega(R/c_p(\omega)\pm t)}}{R}e^{-\alpha(\omega)R}$$
$$=A\sqrt{\rho}\frac{e^{i\omega(R/c_p(\omega)\pm t)}}{R}e^{-\alpha(\omega)R}. \tag{17}$$

This expression is accurate for $\mu\omega\ll 1$ and $\omega\gg cg/R_a T$. The former requirement is accurate for infrasound frequencies and atmospheric viscosities. The latter requirement is valid at periods much less than 5 min, at which point buoyancy effects (gravity waves) are significant. Since the vertical pressure and density gradients depend on gravity, this may suggest that gravitational forces are significant in computing infrasound propagation. However, tests showed that inclusion of gravity terms within the equations governing linear acoustic propagation (e.g., Gill, 1982) led to waveforms that differed by less than 1% over propagation distances of 10 km at frequencies greater than 0.02 Hz. The effect of the density gradient on the wavenumber and velocity becomes significant at lower frequencies than are considered in this paper.

Comparison of Eqs. (14) and (16) indicates that pressure fluctuations in a medium with exponentially decreasing density attenuate more rapidly by a factor of $\sqrt{\rho}$ as they propagate upward, as compared to pressure fluctuations from a source in a uniform density medium. Conversely, the pressure amplitudes decrease less rapidly as the wave propagates downward into denser atmosphere. This follows from consideration of the energy density of the propagating wave (Gill, 1982). Note that that this is not an absorption effect since energy is not lost to the atmosphere.

C. FDTD equations

Though Eqs. (11) correctly describe the physics of acoustic propagation through an attenuating atmosphere, they are not in an optimal form for FDTD implementation. More conveniently, the first-order differential equations in time may be expressed as

$$\frac{\partial\mathbf{v}_s}{\partial t}=\mathbf{f}_i-\frac{1}{\rho}\left(1+\mu\frac{\partial}{\partial t}\right)\nabla p_s, \tag{18a}$$

and

$$\frac{\partial p_s}{\partial t}=-\rho c^2\nabla\cdot\mathbf{v}_s-\xi_s p. \tag{18b}$$

Combining these equations and omitting external forces yields the governing equation Eq. (12), assuming terms involving products of viscosity and density gradients are negligible. Note that, taken individually, Eqs. (18a) and (18b) are not equivalent to Eqs. (11), but combined, each set yields Eq. (12) to first order.

The FDTD approach is based on replacing the spatial and time derivatives in the governing equations by their difference approximations. The system of Eqs. (18a) and (18b) is solved here for two-dimensional (2D) models using a staggered-grid finite-difference technique originally proposed by Yee (1966). The model is divided into a grid of cells of dimension Δx by Δy, each with uniform density and velocity. The pressure variables are sampled at the center of each grid at times $n\Delta t$, where Δt is the temporal step size, and the velocity variables are sampled at points straddling the grid boundaries at times $(n+1/2)\Delta t$. Sampling points in this way yields discretized difference operators that are accurate to second order.

Adopting the notation

$$p_{I,J}^n=p_s(I\Delta x,J\Delta z,n\Delta t), \tag{19a}$$

$$v_{x,I,J}^{n+1/2}=v_{s,x}((I-1/2)\Delta x,J\Delta y,(n+1/2)\Delta t), \tag{19b}$$

$$v_{y,I,J}^{n+1/2}=v_{s,y}(I\Delta x,(J-1/2)\Delta y,(n+1/2)\Delta t), \tag{19c}$$

the set of ions Eqs. (18a) may be replaced by

$$\frac{v_{x,I,J}^{n+1/2}-v_{x,I,J}^{n-1/2}}{\Delta t}=-\frac{1}{\rho_{I-1/2,J}}\left[\left(1+\frac{\mu_{I-1/2,J}}{\Delta t}\right)\left(\frac{p_{I,J}^n-p_{I-1,J}^n}{\Delta x}\right)-\frac{\mu_{I-1/2,J}}{\Delta t}\left(\frac{p_{I,J}^{n-1}-p_{I-1,J}^{n-1}}{\Delta x}\right)\right], \tag{20a}$$

$$\frac{v_{y,I,J}^{n+1/2}-v_{y,I,J}^{n-1/2}}{\Delta t}=-\frac{1}{\rho_{I,J-1/2}}\left[\left(1+\frac{\mu_{I,J-1/2}}{\Delta t}\right)\left(\frac{p_{I,J}^n-p_{I,J-1}^n}{\Delta y}\right)-\frac{\mu_{I,J-1/2}}{\Delta t}\left(\frac{p_{I,J}^{n-1}-p_{I,J-1}^{n-1}}{\Delta y}\right)\right], \tag{20b}$$

where $\rho_{I,J}$ denotes the density, and $\mu_{I,J}$ the shear viscosity values at the center of the $[I,J]$th cell, and the half-integer subscripts denote the average values between two adjacent cells. The discretized version of Eq. (18b) is given by

$$\frac{p_{I,J}^{n+1}-p_{I,J}^{n}}{\Delta t}=-\rho_{I,J}c_{I,J}^{2}\left[\frac{v_{x,I+1,J}^{n+1/2}-v_{x,I,J}^{n+1/2}}{\Delta x}+\frac{v_{y,I,J+1}^{n+1/2}-v_{y,I,J}^{n+1/2}}{\Delta y}\right]-\xi_{I,J}\left[\frac{p_{I,J}^{n+1}+p_{I,J}^{n}}{2}\right]. \tag{21}$$

where $c_{I,J}$ and $\xi_{I,J}$ denote sound speeds and damping values within the $[I,J]$th cell. Equations (20a) and (20b) indicate that only one set of pressure variables p^n must be stored if the μ variable is equal to zero. Introduction of this attenuation coefficient requires saving an extra matrix of the pressure variables, i.e., p^{n-1}, at each time step. The reason for replacing Eqs. (11) with Eqs. (18a) and (18b) is now apparent; use of Eqs. (11) would require saving both the v_x^{n-1} and v_y^{n-1} matrices and would involve may more computational steps, thus increasing the computational load.

The term involving μ in Eq. (18a) is expressed as a backward difference operator in time in Eqs. (20), rather than a central difference operator; thus, it is accurate only to first order. The numerical errors resulting from this choice are examined more closely in the next subsection. The ξp term of Eq. (18b) is computed in Eq. (21) using the time average of the pressure terms at times n and $n+1$. The time derivative on the left-hand side of Eq. (21) is therefore correctly centered at time $n+1/2$.

Equation (21) may be rearranged as

$$p_{I,J}^{n+1}\left(1+\frac{\xi_{I,J}\Delta t}{2}\right)=p_{I,J}^{n}\left(1-\frac{\xi_{I,J}\Delta t}{2}\right)-\rho_{I,J}c_{I,J}^{2}\Delta t\left[\frac{v_{x,I+1,J}^{n+1/2}-v_{x,I,J}^{n+1/2}}{\Delta x}+\frac{v_{y,I,J+1}^{n+1/2}-v_{y,I,J}^{n+1/2}}{\Delta y}\right]. \tag{22}$$

The factor $(1\pm\xi\Delta t/2)$ in the above equation is a first-order Taylor approximation to $\exp[\pm\xi\Delta t/2]$. The update equations for the pressure p may then be expressed more accurately as

$$p_{I,J}^{n+1}=p_{I,J}^{n}e^{-\xi_{I,J}\Delta t}-\rho_{I,J}c_{I,J}^{2}\Delta t\left[\frac{v_{x_{I+1,J}}^{n+1/2}-v_{x_{I,J}}^{n+1/2}}{\Delta x}+\frac{v_{y_{I,J+1}}^{n+1/2}-v_{y_{I,J}}^{n+1/2}}{\Delta y}\right]e^{-\xi_{I,J}\Delta t/2}. \tag{23}$$

Equations (20a) and (20b), and Eq. (23) form the set of coupled equations used in the numerical implementation of the FDTD method.

D. Numerical dispersion and stability

Approximations made in discretizing the continuous partial differential equations lead to numerical dispersion in FDTD solutions (Taflove and Hagness, 2000). The degree of dispersion depends on the ratio of the propagating wavelength to the grid discretization. A rule of thumb for nonattenuating media is that there should be at least 10 nodes per wavelength for the staggered-grid FD method developed by Yee (1966). To compute numerical dispersion for the coupled equations Eqs. (20) and (23), one may consider an homogeneous medium with uniform density, sound speed, viscosity, and damping values. A trial FDTD set of solutions for the velocity and pressure fields of the form

$$V_{x,I,J}^{n+1/2}=V_{x0}e^{i(\omega(n+1/2)\Delta t-\tilde{k}_x(I-1/2)\Delta x-\tilde{k}_yJ\Delta y)}, \tag{24a}$$

$$V_{y,I,J}^{n+1/2}=V_{y0}e^{i(\omega(n+1/2)\Delta t-\tilde{k}_xI\Delta x-\tilde{k}_y(J-1/2)\Delta y)}, \tag{24b}$$

$$P_{I,J}^{n}=P_0e^{i(\omega n\Delta t-\tilde{k}_xI\Delta x-\tilde{k}_yJ\Delta y)}, \tag{24c}$$

may be postulated, where $\tilde{k}_x$ and $\tilde{k}_y$ are the numerical approximations to the horizontal and vertical wavenumbers.

Substitution of the trial wave solutions Eqs. (24a) and (24c) into Eq. (20a) yields the following relation:

$$V_{x0}=P_0\frac{\Delta t}{\rho\Delta x}\frac{\sin(\tilde{k}_x\Delta x/2)}{\sin(\omega\Delta t/2)}\left[1+\frac{\mu}{\Delta t}(1-e^{-i\omega\Delta t})\right]. \tag{25a}$$

Similarly, substituting Eqs. (24b) and (24c) into Eq. (20b) yields

$$V_{y0}=P_0\frac{\Delta t}{\rho\Delta y}\frac{\sin(\tilde{k}_y\Delta y/2)}{\sin(\omega\Delta t/2)}\left[1+\frac{\mu}{\Delta t}(1-e^{-i\omega\Delta t})\right]. \tag{25b}$$

Finally, the trial solutions for the partical velocities may be substituted into Eq. (23) to get

$$P_0\left[\frac{\sin(\omega\Delta t/2-i\xi\Delta t/2)}{\Delta t}\right]=\rho c^2\left[V_{x0}\frac{\sin(\tilde{k}_x\Delta x/2)}{\Delta x}+V_{y0}\frac{\sin(\tilde{k}_y\Delta y/2)}{\Delta y}\right]. \tag{26}$$

Equations (25a) and (25b) are substituted into Eq. (26) to yield a numerical dispersion relation. For the case where a square cell grid is used, i.e., $\Delta x=\Delta y=\Delta$, the numerical dispersion relation may be expressed as

$$\left(\frac{\Delta}{c\Delta t}\right)^2\left[\frac{\sin[(\omega-i\xi)\Delta t/2]\sin(\omega\Delta t/2)}{1+(\mu/\Delta t)(1-e^{-i\omega\Delta t})}\right]=\sin^2\left(\frac{\tilde{k}_x\Delta}{2}\right)+\sin^2\left(\frac{\tilde{k}_y\Delta}{2}\right). \tag{27}$$

The accuracy of the FDTD computations for a given time discretization Δt and grid spacing Δ can be determined using this equation. Defining the Courant stability factor $S=(c\Delta t/\Delta)$ and the grid sampling rate $N_\lambda=(\lambda/\Delta)$, the dispersion relation Eq. (27) may be reformulated as

$$\frac{1}{S^2}\left[\frac{\sin(\pi S/N_\lambda-i\xi\Delta t/2)\sin(\pi S/N_\lambda)}{1+(\mu/\Delta t)(1-e^{-i2\pi S/N_\lambda})}\right]=\sin^2\left(\frac{\Delta\tilde{k}\cos\theta}{2}\right)+\sin^2\left(\frac{\Delta\tilde{k}\sin\theta}{2}\right), \tag{28}$$

where θ is the propagation direction with respect to the x-axis. For $\mu=\xi=0$, this equation reduces to the numerical dispersion relation found in Taflove and Hagness (2000). Equation (28) shows that the numerical dispersion introduced by FD discretization is a function of the propagation

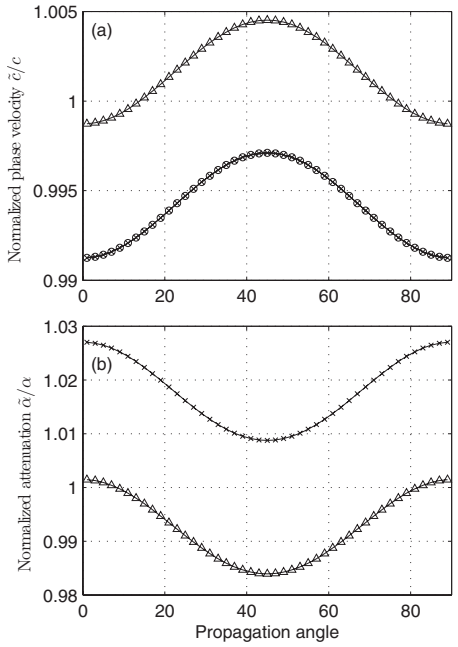


FIG. 4. (a) Variation of the numerical sound-speed ratio $\tilde{c}/c$ with propagation angle for models with no attenuation (circles); $\xi=0.01$, $\mu=0$ (x's); $\xi=0$, $\mu=0.005$ (triangles). The case $\xi=0.01$, $\mu=0$, produces negligibly more numerical dispersion than for the case of no attenuation. Propagation angles are with respect to the horizontal axis. (b) Variation of the attenuation ratios $\mathrm{Im}(\tilde{k})/\mathrm{Im}(k)$ for the models with nonzero attenuation as given above. $S=0.5$, $N_\lambda=12$, and $\Delta t=0.0167$ s were used for each case.

direction as well as the temporal and spatial sampling intervals.

For accurate FDTD simulation the grid sampling rate N_λ, and hence the time sampling rate, is based on the highest frequencies propagated through the grid. As an example, consider a model with sound speed $c=300$ m/s and a source bandwidth of up to 2.5 Hz. Setting $N_\lambda=12$ at the upper end of the frequency range yields a grid spacing of $\Delta=10$ m. A Courant stability factor $S=0.5$ yields a time sampling interval of $\Delta t=0.0167$ s. Figure 4(a) shows the normalized numerical sound speed $\tilde{c}/c=\omega_{\max}/\mathrm{Re}(\tilde{k})$ as a function of propagation angle for models with no attenuation (circles); $\xi=0.01$, $\mu=0$ (x's); $\xi=0$, $\mu=0.005$ (triangles). Figure 4(b) shows the normalized attenuation factors $\mathrm{Im}(-\tilde{k})/(\xi/(2c))$ and $\mathrm{Im}(-\tilde{k})/(\mu*\omega^2/(2c))$ for the latter two models. Note that the numerical phase velocity computed here differs from the dispersive velocities examined in Sec. III as it results from the discretization of the wave equations. Only in the case $\lim \Delta t\to 0$ and $\lim \Delta\to 0$, it can be shown that the numerical wavenumber $\tilde{k}^2$ approaches k^2 as defined by Eq. (4).

As indicated in Fig. 4, both the numerical phase velocity and attenuation depend weakly on the propagation angle; however, the difference is less than 2% for the given discretization values. Furthermore, variations in numerical phase velocities exists whether or not attenuation is included. For $\mu=0$, $\xi\neq 0$ the ξ value has minimal effect on the numerical phase velocity and attenuation over frequencies and attenuation values relevant to global infrasound propagation. However, for $\xi=0$, $\mu\neq 0$, the accuracy of the FDTD solution

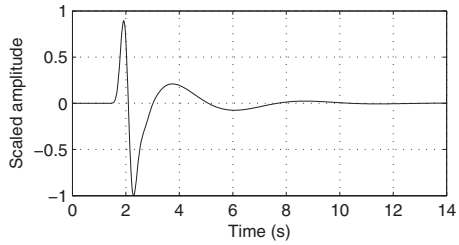


FIG. 5. The source function used for examples A, B, and C. This source was generated by bandpass filtering a spikelike waveform between 0.02 and 2.2 Hz.

degrades with increasing μ values—the normalized velocities increase and the normalized attenuation factors decrease.

Finally, it is known that for a 2D model the time step Δt must be chosen such that $S<1/\sqrt{2}$ to ensure stability (Taflove and Hagness, 2000). Neglecting ξ, it can be shown that for $\omega_{\max}\mu\ll 1$, the stability requirement is given by

$$\Delta t<\frac{\Delta}{\sqrt{2}c(1+\omega_{\max}^2\mu^2/4)}. \tag{29}$$

The conditions $\omega_{\max}\mu\ll 1$ and Eq. (29) provide stability limits on the computations. Thus, the requirement $\omega_{\max}\mu\ll 1$ imposes an upper limit on μ or, conversely, on $\omega_{\max}$ for FDTD computations.

V. EXAMPLES

Several FDTD solutions to problems of infrasound propagation through an absorbing atmosphere are presented in this section. Examples A and B are idealized problems used to test and illustrate the effects of including attenuation coefficients in the FDTD method. Example C involves a more realistic atmospheric waveguide. For each example the side and top boundaries are terminated using a perfectly matched layer (PML) absorbing boundary condition (Berenger, 1994) initially developed for electromagnetic waves. The PML boundary conditions are also applied to the bottom boundary for examples A and B, that is, the models are assumed to be whole spaces. The source function shown in Fig. 5, which consists of a high-frequency pulse followed by a lower-frequency wave, was used for each example.

Example A consists of a uniform whole space with constant density and a constant sound speed of 330 m/s. Pressure waves are sampled at distances from 2 to 10 km from the source, at intervals of 2 km. The node spacing is 10 m, yielding 15 nodes per wavelength at the maximum source frequency of 2.2 Hz, and the time sampling interval is 0.0168 s, corresponding to $S=1/1.8$. Results were synthesized for a model with no attenuation and compared to three attenuation models: first, a model with constant attenuation values of $\xi=0$, $\mu=0.002$; from Eq. (10a), this corresponds to an attenuation coefficient of $1.039\times 10^{-3}\ f^2$ dB/m; second, a model with constant attenuation values of $\xi=0$, $\mu=0.005$, corresponding to an attenuation coefficient of $2.60\times 10^{-3}\ f^2$ dB/m; and third, a model with $\xi=0.01$, $\mu=0$; from Eq. (10b), this corresponds to an attenuation coefficient of 1.316×10^{-4} dB/m.

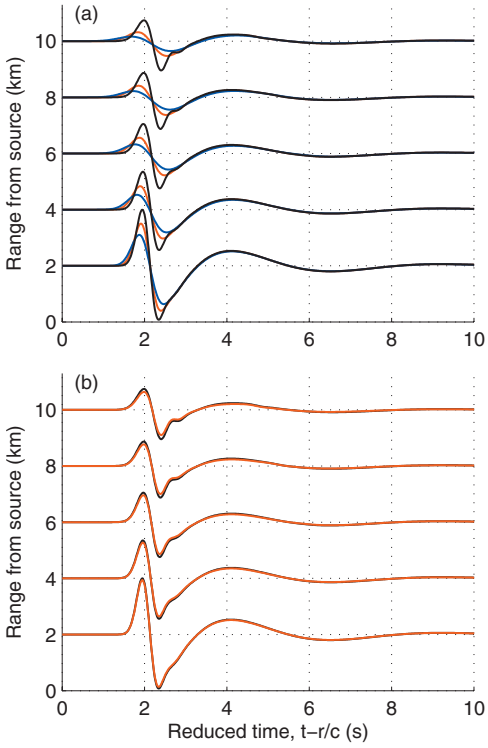


FIG. 6. (a) Synthesized waveforms for models with varying μ values at distances from 2 to 10 km from the source. The pressure waveforms are shown as a function of reduced time, which is the time minus the range divided by the sound speed (c=330 m/s). The waveforms for the models with no attenuation are shown by the black line; the results for μ=0.002 are indicated in red, and the results for μ=0.005 are indicated in blue. (b) The black line shows the pressure solutions for the model with no attenuation; the red lines indicates waveforms corresponding to models with ξ=0.01.

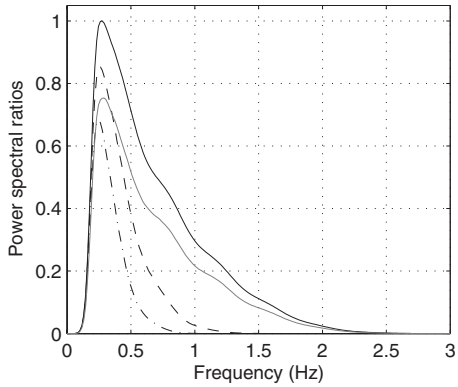


FIG. 7. Ratios of power spectra for pressure waveforms at a distance of 10 km from the source for example A. Power spectra are normalized with respect to the peak power of the waveform for the model with no attenuation (solid black line). The dashed line corresponds to the ξ=0, μ=0.002 model, the dash-dot line corresponds to the ξ=0, μ=0.005 model, and the gray line corresponds to the ξ=0.01, μ=0 model.

The synthesized waveforms for models with varying μ are shown in Fig. 6(a). The waveforms for models with no attenuation show two-dimensional geometrical spreading away from the source. In comparison, waveforms propagated through models with nonzero μ values show greater attenuation of the initial higher frequency pulse, but the later, low-frequency wave at approximately 4 s is virtually unchanged. Furthermore, the waveforms for the nonzero μ models are slightly advanced with respect to the nonattenuated waveforms. The numerical dispersion relation, Eq. (28), predicts that the phase velocity for the model with μ=0.005 should be about 0.5% faster than the model with no attenuation, as shown in Fig. 4(a), and in agreement with Eq. (7a), which also predicts higher phase velocities with increasing μ. Synthesized waveforms for models with no attenuation are almost identical to those for the model with ξ=0.01, as shown in Fig. 6(b). This is approximately the highest value of the constant attenuation coefficient ξ realized in the Earth's atmosphere, as indicated in Fig. 2. As shown in Fig. 6(b), the constant attenuation term conserves the shape of the waveform.

The power spectra for the pressure waveforms sampled at a distance of 10 km from the source are shown in Fig. 7. As indicated, waveforms for models with $\mu \neq 0$ are most significantly attenuated at the highest frequencies sampled, whereas waveforms for models with $\xi \neq 0$ are uniformly attenuated at all frequencies. For ξ=0.01 the power spectral ratio of attenuated to unattenuated waveforms at 10 km is 0.736 for the entire source bandwidth, corresponding to a transmission loss of 1.33 dB; this agrees to within 1% with the predicted transmission loss of 1.32 dB at 10 km for this value. ξ value. For the μ=0.002 model, the attenuation at a frequency of 0.25 Hz is 0.629 dB, which compares well with the predicted transmission loss of 0.631 dB at this distance and frequency. Similarly, the attenuation for the μ=0.005 model, observed to be 1.57 dB at 0.25 Hz and 10 km from the source, compares well with the predicted transmission loss of 1.58 dB.

The model for example B is similar to example A with the difference that the atmospheric density decreases exponentially with altitude at a rate of three orders of magnitude over 50 km. Pressure waves are sampled at altitudes from 2 to 8 km both above and below the source, at intervals of 2 km. Waveforms were synthesized twice; once for a model with no attenuation and once for a model with ξ=0, μ=0.002, as shown in Fig. 8. Comparing this to Fig. 6(a), it may be seen that waveform amplitudes decrease more rapidly with increasing distance from the source as they propagate upward into thinner atmosphere than for a model with constant density. Conversely, the waveform's amplitudes decrease much more slowly as the acoustic energy propagates downward. Again, including a nonzero μ term in the model leads to waveforms with suppressed high-frequency pulses.

Normalized power spectra for each waveform are shown in Fig. 9. The power spectra corresponding to the model with μ=0.002 lie along diminishing curves as the distance from the source increases. The transmission loss values of 0.148, 0.284, 0.428, and 0.576 dB for the synthesized waveforms at the peak frequency of 0.26 Hz compare well with the predicted transmission loss values of 0.142, 0.283, 0.425, and 0.566 dB at this frequency.

Example C is similar to example B except that both realistic sound-speed and attenuation profiles for an altitude range of 89 to 111 km are used, as shown in Fig. 2. The source is located at an altitude of 100 km and waveforms are sampled at altitudes from 2 to 8 km above and below the source, at intervals of 2 km. Since the sound speeds are

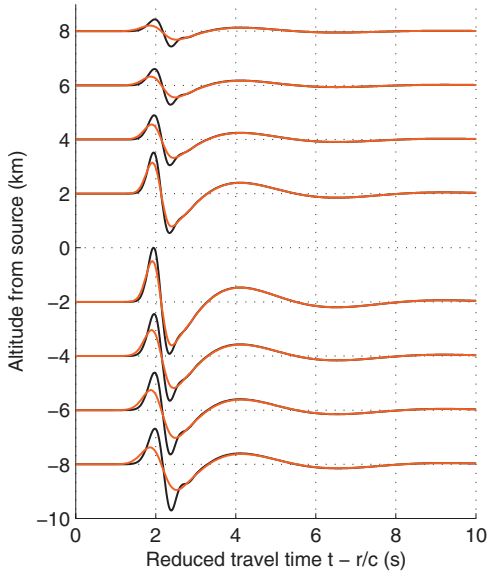


FIG. 8. Synthesized pressure waveforms for example B at distances from 2 to 8 km above (upper waveforms) and below the source located at 0 km (lower waveforms), shown as a function of reduced time. Results for a model with no attenuation are shown in black and results for a model with $\mu=0.002$ are shown in red.

lower for this final example, the model was discretized in cells of 8 m $\times$ 8 m and the time sampling interval was set to 0.0135 s. Waveforms were synthesized both with and without attenuation. Results are shown in Fig. 10. Arrivals are earlier for upward propagation due to variations in the sound-speed profiles, as shown in Fig. 2.

As indicated in Fig. 10, the high frequencies are more severely attenuated as they propagate upward from the source rather than downward, since the atmospheric absorption increases with altitude. This is seen more clearly in Fig. 11, where the normalized power spectra for each attenuated waveform are compared to the corresponding nonattenuated

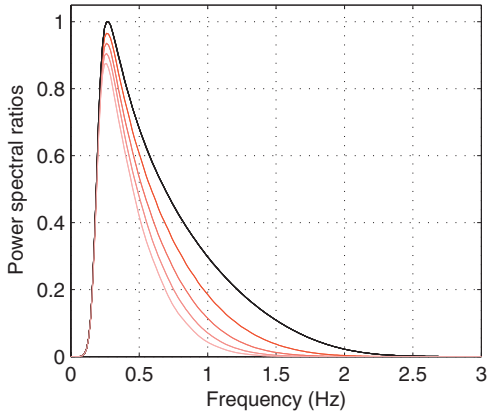


FIG. 9. Normalized power spectra for each waveform shown in Fig. 8. At each range step, the power spectra are normalized with respect to the peak power for waveforms propagated through the nonabsorbing medium. Power spectra for the waveforms synthesized from the model with no attenuation lie along a single curve, shown by the heavy line, indicating that spectral content is preserved. The red lines, which become lighter with increasing distance from the source, are the normalized power spectra for waveforms at distances of 2, 4, 6, and 8 km from the source for a model with $\mu=0.002$. The red lines show that high frequencies are diminished with increasing distance from the source.

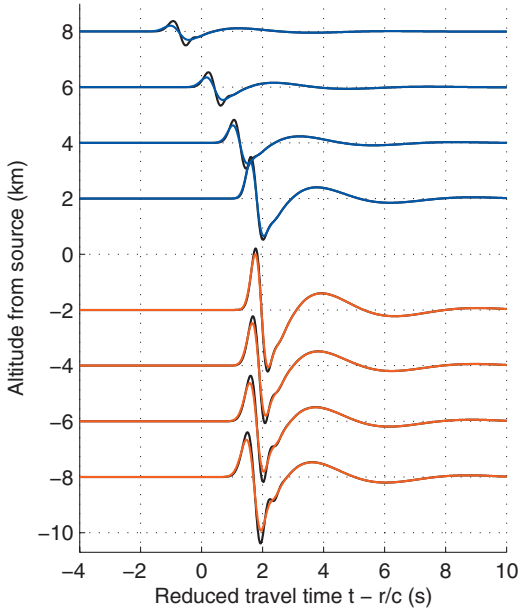


FIG. 10. Synthesized pressure waveforms for example C at distances from 2 to 8 km above and below the source located at 100 km. Results for a model with no attenuation are shown in black and results for a model with realistic attenuation profiles for this altitude range are shown in red for down-going waves and blue for up-going waves. The reducing velocity was set to 260 m/s.

waveforms. As in Example B, the power spectra for waveforms lie along diminishing curves with increasing distance from the source, but here the attenuation is greater for the receivers above the source.

VI. DISCUSSION AND CONCLUSIONS

A FDTD technique has been developed for acoustic propagation through a heterogeneous absorbing atmosphere. The attenuation terms used in the propagation modeling were defined in such a way as to match observed variations in absorption with frequency. It was demonstrated that, for fre-

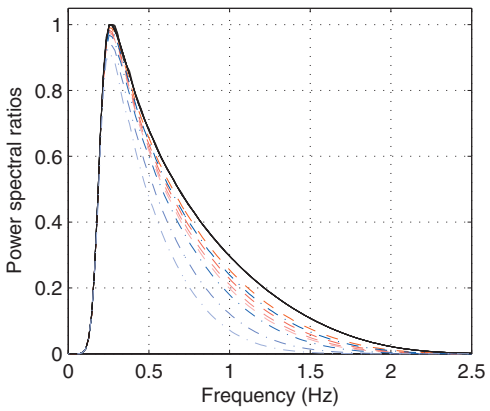


FIG. 11. Normalized power spectra for each waveform shown in Fig. 10. At each range step, the power spectra are normalized with respect to the peak power for waveforms propagated through a nonabsorbing medium. Power spectra for waveforms synthesized from the model with no attenuation lie along a single curve, shown by the black heavy line. The red dashed lines are the corresponding power spectra for down-going waves at distances of 2, 4, 6, and 8 km from the source and the blue dash-dotted lines are for up-going waves. Waveforms undergo greater attenuation for upward propagation than for downward propagation.

quency bandwidths of approximately two orders of magnitude, infrasound attenuation within the atmosphere may be modeled as the sum of a constant term β, and another term γ multiplied by the square of the frequency. For infrasound propagation within Earth's atmosphere, the former term dominates at altitudes of approximately 30–50 km, i.e., within the upper stratosphere; the μ term dominates at other altitudes.

Inclusion of attenuation terms in the wave equation is associated with some velocity dispersion, i.e., the sound velocity becomes a function of frequency. However, this effect is small except at very high attenuation levels within the upper atmosphere, and at the upper end of the frequency range of interest.

The accuracy of the absorbing FDTD technique was tested for several problems involving models with uniform velocity. By comparing results for models with and without attenuation terms, it was shown that the predicted transmission losses for the attenuating media agreed with those computed from the synthesized waveforms. Inclusion of the constant attenuation term does not reduce the stability or computation time of the algorithm. However, inclusion of the $\mu \neq 0$ term increases the computation time by approximately 50% and requires storage of an extra pressure time step. It also imposes extra conditions on the product of the frequency and viscosity, $\omega_{\max} \mu_{\max} \ll 1$, and on the temporal sampling interval $\Delta t < \Delta / [\sqrt{2} c (1 + \omega_{\max}^2 \mu^2 / 4)]$; these conditions are required to maintain stability of the two-dimensional FDTD algorithm. Note that the best-fit $\mu(z)$ profile depends on the frequency bandwidth of interest and decreases at higher frequency ranges; thus, this requirement does not preclude the use of this algorithm at higher frequencies.

Possible extensions of this work involve including terms in the wave equations to account for wind and gravitational effects. Wind affects infrasound propagation at all frequencies; gravitational effects could alter the waveform solutions at lower frequencies than are of interest here. Other extensions include testing the accuracy of the Sutherland and Bass (2004) models by comparing numerically synthesized waveforms derived from the algorithm presented here against experimental observations.

ACKNOWLEDGMENTS

Thanks to Claus Hetzer and Hank Bass of the National Center of Physical Acoustics at the University of Mississippi for providing their code to compute attenuation coefficients as a function of frequency and altitude. This work was supported by UM Subcontract No. 07-08-013. Distribution A. Approved for Public Release; distribution unlimited. Finally, thanks to three anonymous reviewers for their careful reviews. Their recommendations for improving the manuscript are sincerely appreciated.

APPENDIX: ACOUSTIC PROPAGATION IN A VISCOUS FLUID

The equations governing linear acoustic propagation through an attenuating fluid or gas may be derived as outlined in this Appendix.

The equation of motion relating the fluid acceleration to the stress tensor σ_{ij} and the external volume force per unit mass f_i is given by

$$\rho \frac{Dv_i}{Dt} = \rho f_i + \frac{\partial \sigma_{ij}}{\partial x_j}, \tag{A1}$$

(Eq. 3.2.2 of Batchelor, 1967) where ρ denotes the density, v_i is the particle velocity, and D/Dt is the advective derivative, $D/Dt = \partial/\partial t + \mathbf{v} \cdot \boldsymbol{\nabla}$.

A Newtonian fluid is one in which the shear stress is linearly proportional to the rate of shear deformation. In this case the stress-strain relationship in summation notation is given by

$$\sigma_{ij} = -p \delta_{ij} + A_{ijkl} \frac{\partial v_k}{\partial x_l}, \tag{A2}$$

(Eq. 7.3-1 of Fung, 1977) where p is the mean normal stress, δ_{ij} is the Kronecker delta, equal to 1 if $i=j$ and 0 otherwise, and the fourth-order tensor coefficients A_{ijkl} depend on the state of the fluid, i.e., the temperature, but not on the stress or rate of deformation. The second term on the right represents the deviatoric stress tensor and is defined such that the diagonal terms sum to zero. The stress tensor may be expressed as

$$\sigma_{ij} = -p \delta_{ij} + 2\mu' (e_{ij} - 1/3 \boldsymbol{\Delta} \delta_{ij}), \tag{A3}$$

if the fluid is isotropic (Eq. 3.2.11 of Batchelor, 1967; Eq. 7.3.6 of Fung, 1977), where μ' is the viscosity in units of kg m^{-1} s^{-1}, $e_{ij} = 1/2(\partial v_i/\partial x_j + \partial v_j/\partial x_i)$, and $\boldsymbol{\Delta}$ is the rate of expansion given by $\boldsymbol{\Delta} = \boldsymbol{\nabla} \cdot \mathbf{v} = e_{kk}$. A fluid with behavior that obeys Eq. (A3) is called a Stokes' fluid.

Substitution of Eq. (A3) for the stress tensor into Eq. (A1) yields the following equation of motion for a Stokes' fluid:

$$\rho \frac{Dv_i}{Dt} = \rho f_i - \frac{\partial p}{\partial x_i} + \mu' \left(\boldsymbol{\nabla}^2 v_i + \frac{1}{3} \frac{\partial \boldsymbol{\Delta}}{\partial x_i} \right). \tag{A4}$$

As discussed in Batchelor (1967), Eqs. (A3), and hence Eq. (A4), were derived under the assumption that the rate of isotropic expansion does not contribute to the stress; thus, the mean normal stress p is equal to the pressure. If a damping term is added the equation of motion as an external force, Eq. (A4) becomes

$$\rho \frac{Dv_i}{Dt} = \rho f_i - \frac{\partial p}{\partial x_i} + \mu' \left(\boldsymbol{\nabla}^2 v_i + \frac{1}{3} \frac{\partial \boldsymbol{\Delta}}{\partial x_i} \right) - \xi' v_i, \tag{A5}$$

where ξ' is the damping coefficient in units of kg m^{-3} s^{-1}. The inclusion of this term is entirely heuristic and is used to provide an additional degree of freedom to the governing equations and thus provide a better fit with observed attenuation values, particularly at midaltitudes. Gravitational effects may also be included as an external force but are ne-

glected here as they are relevant to much lower frequencies than are dealt with in this paper.

In the presence of fluctuations caused by the passage of a sound wave, the total pressure, density, and velocity may be expressed as $p=p_o+p_s$, $v=w+v_s$, and $\rho=\rho_o+\rho_s$, where p_o, w, and ρ_o are ambient values, and p_s, v_s, and ρ_s are small fluctuations caused by the sound wave. A linear expression for the particle velocity in the presence of a sound wave may be derived by inserting these values into Eq. (A5), and retaining only first-order terms in the pressure, velocity, and density terms. Furthermore, the effects of wind are neglected in this paper. Then,

$$\frac{\partial v_{s,i}}{\partial t}=f_i-\frac{1}{\rho_o}\left[\frac{\partial p_s}{\partial x_i}-\mu'\left(\nabla^2 v_{s,i}+\frac{1}{3}\frac{\partial \Delta_s}{\partial x_i}\right)+\xi' v_{s,i}\right], \quad \text{(A6)}$$

where $\Delta_s=\nabla\cdot\mathbf{v}_s$. In vector notation, this may be written as

$$\frac{\partial \mathbf{v}_s}{\partial t}=\mathbf{f}_i-\frac{1}{\rho_o}\left[\nabla p_s-\mu'\left(\nabla^2\mathbf{v}_s+\frac{1}{3}\nabla(\nabla\cdot\mathbf{v}_s)\right)+\xi'\mathbf{v}_s\right], \quad \text{(A7)}$$

where $\nabla^2\mathbf{v}_s$ is a vector with components $\nabla^2\mathbf{v}_s=(\nabla^2\mathbf{v}_{s,x},\nabla^2\mathbf{v}_{s,y},\nabla^2\mathbf{v}_{s,z})$. Viscosity values μ' are nearly constant within the atmosphere up to an altitude of 86 km (Gill, 1982).

The equation of state for air is given by $p=p(\rho,S)$, where S is the entropy. Applying the advective derivative to the equation of state yields

$$\frac{Dp}{Dt}=\tilde{c}^2\frac{D\rho}{Dt}+h\frac{DS}{Dt}, \quad \text{(A8)}$$

where $\tilde{c}^2=\partial p/\partial\rho$ is the squared sound speed in the presence of a sound wave, and $h=\partial p/\partial S$. The first term on the right denotes the change in pressure associated with a density change. The second term represents the increase in entropy caused by the passage of a sound wave, and is associated with an increase in temperature due to viscous dissipation. In the absence of a mass source, the conservation of mass equation is given by

$$\frac{D\rho}{Dt}=-\rho\nabla\cdot\mathbf{v}, \quad \text{(A9)}$$

so the first term on the right of Eq. (A8) is given by $-\rho\tilde{c}^2\nabla\cdot\mathbf{v}$.

The entropy per unit mass η is related to the internal energy per unit mass E through

$$Td\eta=dE+pd(1/\rho), \quad \text{(A10)}$$

(Eq. 3.2.1 of Gill, 1982), where T is the temperature. Applying the advective derivative leads to

$$T\frac{D\eta}{Dt}=\frac{DE}{Dt}-\frac{p}{\rho^2}\frac{D\rho}{Dt}. \quad \text{(A11)}$$

Using the equation for the conservation of mass leads to

$$T\frac{D\eta}{Dt}=\frac{DE}{Dt}+\frac{p}{\rho}\nabla\cdot\mathbf{v}. \quad \text{(A12)}$$

From Eq. (3.4.4) of Batchelor (1967),

$$\frac{DE}{Dt}=-\frac{p}{\rho}\nabla\cdot\mathbf{v}+\frac{2\mu}{\rho}\left(e_{ij}e_{ij}-\frac{1}{3}(\nabla\cdot\mathbf{v})^2\right)+\frac{1}{\rho}\frac{\partial}{\partial x_i}\left(k\frac{\partial T}{\partial x_i},\right), \quad \text{(A13)}$$

where the first term on the right is related to compression or expansion, the second term relates to the shear viscosity, and the final term relates to heat exchange by molecular conduction, which is not considered further. Combining Eqs. (A12) and (A13) (neglecting the conduction term) leads to

$$T\frac{D\eta}{Dt}=\frac{2\mu}{\rho}\left(e_{ij}e_{ij}-\frac{1}{3}(\nabla\cdot\mathbf{v})^2\right). \quad \text{(A14)}$$

The change in the entropy per unit mass is proportional to the square of the particle velocity perturbation in the absence of wind, and is thus negligible to first order for linear propagation and for small μ.

Since propagation is isentropic to first order, the second term of Eq. (A8) may be neglected. It may also be shown that for small perturbations $\tilde{c}^2=c^2$ (e.g., Gill, 1982), where $c^2=\partial p_o/\partial\rho_o$ is the squared static sound speed. Then, combining Eqs. (A8) and (A9) and retaining only first-order terms leads to the following equation for pressure fluctuations due to a sound wave:

$$\frac{\partial p_s}{\partial t}=-\rho_o c^2\nabla\cdot\mathbf{v}_s, \quad \text{(A15)}$$

where pressure gradients have been neglected. Equations (A7) and (A15) form a complete set of equations describing linear acoustic propagation through an attenuating fluid or gas.

Applying the operator $\partial/\partial t$ to Eq. (A15) and making use of Eq. (A7) yields the second-order equation for p_s,

$$\frac{\partial^2 p_s}{\partial t^2}=c^2[\nabla^2 p_s-4\mu'/3\nabla^2(\nabla\cdot v_s)+\xi'(\nabla\cdot v_s)]-\frac{c^2}{\rho_o}\nabla\rho_o\cdot\nabla p_s+O\left(\rho_o\nabla\left(\frac{\mu'}{\rho_o}\right)\right)+O\left(\rho_o\nabla\left(\frac{\xi'}{\rho_o}\right)\right) \quad \text{(A16)}$$

where the external force f_i is omitted, and the identity $\nabla\cdot\nabla^2\mathbf{v}_s=\nabla^2(\nabla\cdot\mathbf{v}_s)$ has been used. The final two terms relate to gradients of the viscosity divided by the atmospheric density. These terms may be omitted for realistic atmospheric density and viscosity values. So, making use of Eq. (A15), Eq. (A16) may be expressed as

$$\frac{\partial^2 p_s}{\partial t^2}=c^2\left[1+\mu\frac{\partial}{\partial t}\right]\nabla^2 p_s-\xi\frac{\partial p}{\partial t}-\frac{c^2}{\rho_o}\nabla\rho_o\cdot\nabla p_s, \quad \text{(A17)}$$

where

$$\mu=4\mu'/(3\rho_o c^2);\quad \xi=\xi'/\rho_o, \quad \text{(A18)}$$

so that μ is in units of seconds, and ξ is in units of s^{-1}. Equation (A17) is the governing equation for acoustic propagation through an attenuating atmosphere, although it may be

noted that the inclusion of the term involving ξ was heuristic, used only to yield more realistic atmospheric absorption values.

Aki, K., and Richards, P. G. (**1980**). *Quantitative Seismology: Theory and Methods*, (W. H. Freeman and Co., San Francisco), Vol. **1**.

Arrowsmith, S. J., and Hedlin, M. A. H. (**2005**). "Observations of infrasound from surf in southern California," Geophys. Res. Lett. **32**, L09810.

Bass, H. E., Sutherland, L. C., Zuckerwar, A. J., Blackstock, D. T., and Hester, D. M. (**1995**). "Atmospheric absorption of sound: Further developments," J. Acoust. Soc. Am. **97**, 680–683.

Bass, H. E., Hetzer, C. H., and Raspet, R. (**2007**). "On the speed of sound in the atmosphere as a function of altitude and frequency," J. Geophys. Res. **112**, D15110.

Bass, H., Bhattacharyya, J., Garces, M., Hedlin, M. A. H., Olson, J., and Woodward, R. (**2006**). "Infrasound," Acoust. Today **2**, 9–19.

Batchelor, G. K. (**1967**). *An Introduction to Fluid Dynamics* (University Printing House, Cambridge).

Berenger, J. P. (**1994**). "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 1185–1200.

Blanch, J. O., Robertsson, J. O. A., and Symes, W. W. (**1995**). "Modeling of a constant Q: Method and algorithm for an efficient and optimally inexpensive viscoelastic technique," Geophysics **60**, 176–184.

Botteldoren, D. (**1994**). "Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid," J. Acoust. Soc. Am. **95**, 2213–2219.

Boyce, W. E., and DiPrima, R. C. (**1977**). *Elementary Differential Equations and Boundary Values Problems*, 3rd ed., (Wiley and Sons, Toronto).

Dain, Y., and Lueptow, R. M. (**2001**). "Acoustic attenuation in a three-gas mixture: Results," J. Acoust. Soc. Am. **110**, 2974–2979.

Day, S. M. (**1998**). "Efficient simulation of constant Q using coarse-grained memory variables," Bull. Seismol. Soc. Am. **88**, 1051–1062.

Day, S. M., and Minster, J. B. (**1984**). "Numerical simulation of attenuated wavefields using a Padé approximant method," Geophys. J. R. Astron. Soc. **78**, 105–118.

Der, Z. A., Shumway, R. H., and Herrin, E. T. (**2002**). *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Data Processing and Infrasound*, Pageoph Topical Volumes (Birkhuser, Basel).

Ejakov, S. G., Phillips, S., Dain, Y., Lueptow, R. M., and Visser, J. H. (**2003**). "Acoustic attenuation in gas mixtures with nitrogen: Experimental data and calculations," J. Acoust. Soc. Am. **113**, 1871–1879.

Emmerich, H., and Korn, M. (**1987**). "Incorporation of attenuation into time-domain computations of seismic wave fields," Geophysics **52**, 1252–1264.

Evans, L. B., Bass, H. E., and Sutherland, L. C. (**1972**). "Atmospheric absorption of sound: Theoretical predictions," J. Acoust. Soc. Am. **51**, 1565–1575.

Fung,Y. C. (**1977**). *A First Course in Continuum Mechanics*, 2nd ed. (Prentice-Hall, Inc., Englewood Cliffs, NJ).

Garceś, M., Willis, M., Hetzer, C., and Drob, D. (**2004a**). "On using ocean swells for continuous infrasonic measurements of winds in the lower, middle, and upper atmosphere," Geophys. Res. Lett. **31**, L19304.

Garceś, M., Bass, H., Drob, D., Hetzer, C., Hedlin, M., Le Pichon, A., Lindquist, K., North, R., and Olson, J. (**2004b**). "Forensic studies of infrasound from massive hypersonic sources," EOS (Wash. D.C.) **85** (43), 433–440.

Gill, A. E. (**1982**). *Atmosphere-Ocean Dynamics* (Academic, San Diego).

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (AIP, Woodbury, NY).

Knopoff, L. (**1956**). "The seismic pulse in materials possessing solid friction. I. Plane waves," Bull. Seismol. Soc. Am. **46**, 175–183.

Le Pichon, A., Herry, P., Mialle, P., Vergoz, J., Brachet, N., Garceś, M., Drob, D., and Ceranna, L. (**2005**). "Infrasound associated with 2004–2005 large Sumatra earthquakes and tsunami," Geophys. Res. Lett. **32**, L19802.

Le Pichon, A., Garceś, M., Blanc, E., Barthelémy, M., and Drob, D. P. (**2002**). "Acoustic propagation and atmosphere characteristics derived from infrasonic waves generated by the Concorde," J. Acoust. Soc. Am. **111**, 629–641.

Lighthill, J. (**1978**). *Waves in Fluids* (Cambridge University Press, Cambridge, UK).

Lingevitch, J. F., Collins, M. D., and Siegmann, W. L. (**1999**). "Parabolic equations for gravity and acousto-gravity waves," J. Acoust. Soc. Am. **105**, 3049–3056.

Mikumo, T. (**1968**). "Atmospheric pressure waves and tectonic deformation associated with the Alaskan earthquake of March 28, 1964," J. Geophys. Res. **73**, 2009–2025.

Ostashev, V. E., Wilson, D. K., Liu, L., Aldridge, D. F., Symons, N. P., and Marlin, D. (**2005**). "Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation," J. Acoust. Soc. Am. **117**, 503–517.

ReVelle, D. O., Brown, P. G., and Spurny, P. (**2004**). "Entry dynamics and acoustics/infrasonic/seismic analysis for the Neuschwanstein meteorite fall," Meteorit. Planet. Sci. **39**, 1605–1625.

Sutherland, L. C., and Bass, H. E. (**2004**). "Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **115**, 1012–1032.

Sutherland, L. C., and Bass, H. E. (**2006**). "Erratum: Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **120**, 2985.

Taflove, A., and Hagness, S. C. (**2000**). *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 2nd ed. (Artech House, Norwood, MA).

Wilson, C. R., and Forbes, R. B. (**1969**). "Infrasonic waves from Alaskan volcanic eruption," J. Geophys. Res. **74**, 4511–4522.

Yee, K. S. (**1966**). "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas Propag. **14**, 302–307.

Young, J. M., and Greene, G. E. (**1982**). "Anomalous infrasound generated by the Alaskan earthquake of March 28, 1964," J. Acoust. Soc. Am. **71**, 334–339.

Evaluation of infrasound signals from the shuttle Atlantis using a large seismic network

Catherine D. de Groot-Hedlin, Michael A. H. Hedlin, and Kristoffer T. Walker
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, California 92093-0225

Douglas P. Drob
Space Sciences Division, Naval Research Laboratory, 4555 Overlook Avenue, SW, Washington, DC 20375

Mark A. Zumberge
Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, California 92093-0225

(Received 6 December 2007; revised 10 June 2008; accepted 11 June 2008)

Inclement weather in Florida forced the space shuttle "Atlantis" to land at Edwards Air Force Base in southern California on June 22, 2007, passing near three infrasound stations and several hundred seismic stations in northern Mexico, southern California, and Nevada. The high signal-to-noise ratio, broad receiver coverage, and Atlantis' positional information allow for the testing of infrasound propagation modeling capabilities through the atmosphere to regional distances. Shadow zones and arrival times are predicted by tracing rays that are launched at right angles to the conical shock front surrounding the shuttle through a standard climatological model as well as a global ground to space model. The predictions and observations compare favorably over much of the study area for both atmospheric specifications. To the east of the shuttle trajectory, there were no detections beyond the primary acoustic carpet. Infrasound energy was detected hundreds of kilometers to the west and northwest (NW) of the shuttle trajectory, consistent with the predictions of ducting due to the westward summer-time stratospheric jet. Both atmospheric models predict alternating regions of high and low ensonifications to the NW. However, infrasound energy was detected tens of kilometers beyond the predicted zones of ensonification, possibly due to uncertainties in stratospheric wind speeds.
 [DOI: 10.1121/1.2956475]

PACS number(s): 43.28.Fp, 43.28.Mw, 43.20.Dk [VEO] Pages: 1442–1451

I. INTRODUCTION

A primary goal in infrasound research is to accurately model transmission of low frequency acoustic energy to distances of several hundreds to thousands of kilometers—a challenging task given that the propagation characteristics are a function of the background winds and atmospheric temperatures, which may vary considerably in space and time (Drob *et al.*, 2003). Unlike in seismology, where Earth's velocity structure is largely derived from the travel times of many intersecting seismic waves, the time dependent atmospheric models used for infrasound propagation modeling are not derived from direct recordings of sound waves, so there is no assurance that they will yield predictions that agree with observations. Infrasound studies are thus inextricably linked to the atmospheric sciences, as diverse data sets ranging from direct observations to climatological models provide the specifications of the sound and wind speed profiles required for modeling infrasound propagation. An inability to accurately model a set of infrasound observations indicates a knowledge gap in our understanding and ability to specify the state of the atmosphere. Recently, studies have demonstrated the utility of infrasound research to advance scientific understanding of the atmosphere (see, e.g., Le Pichon *et al.*, 2002; Millet *et al.*, 2007; and Arrowsmith *et al.*, 2007).

The objective of this paper is to use observations to evaluate our ability to model low frequency sound propagation through the atmosphere for a single ground-truth event. This relies on our knowledge of the atmosphere at the time of the event, the infrasound source, as well as the physical assumptions used in propagation modeling. Until recently, there have been limited opportunities to evaluate present-day atmospheric models using infrasound signals from sources for which we know exactly when and where the source occurred. The recent re-entry of the space shuttle Atlantis over the heavily instrumented western United States provided one such opportunity. Although the shuttle usually lands at the Kennedy Space Center (KSC) in Florida, severe weather in the vicinity of KSC on June 22, 2007 forced NASA to direct Atlantis to the alternate landing site at Edwards Air Force Base (AFB) in the Mojave Desert in southern California. A time and position record of its trajectory was recorded with an onboard GPS receiver. On its approach to Edwards AFB, the shuttle passed just west of Baja California and then across San Diego and Los Angeles before passing below the sound barrier just above the Mojave Desert. A double sonic boom was heard by millions of people and, fortunately for this study, was recorded by over 100 three-component seismic stations of the USArray (Meltzer *et al.*, 1999), various regional seismic networks, and three infrasound arrays in

southern California and western Nevada. The temporary presence of the transportable USArray in this region provides this study with a much broader and denser array of sensors than would have otherwise been available. Although seismic wave forms do not yield an accurate measure of the infrasound amplitude or structure due to local variations in the efficiency of air to ground coupling, they can provide both travel times and a lower limit on the ground exposure to the sonic boom.

Previous studies involving seismic recordings of shock waves include using arrival times to estimate a rough trajectory of supersonic objects including a bolide (Ishihara *et al.*, 2004; Langston, 2004) and the space shuttle Columbia (Kanamori *et al.*, 1992). Wave form characteristics have been used to estimate the seismic velocity of near-surface layers (Langston, 2004) and to infer mechanisms by which ground roll is excited (Kanamori *et al.*, 1991, 1992). These studies made use of a constant sound speed to model acoustic propagation through the atmosphere. This approximation is generally adequate for sound propagation to the primary sound carpet, which is associated with downward propagation of sound from supersonic objects to the ground and has a width of tens of kilometers. It can also be accurate to first order at greater distances; Cochran and Shearer (2006) employed a constant acoustic velocity to infer the location of explosive sources recorded at seismic networks in Southern California, and Gibbons *et al.* (2007) applied joint seismic-infrasound processing to arrivals recorded at small seismic arrays to find the direction of arrival of explosive sources.

An accurate description of the atmospheric state is required to model propagation at greater lateral distances from the flight trajectory, where arrivals associated with refraction from the stratosphere may form a secondary sound carpet. Ducting between the ground and temperature and wind speed gradients in the lower, middle, and upper atmosphere allows infrasound waves to propagate to distances of hundreds to thousands of kilometers. Concorde flights have been recorded at infrasound sensors at distances from hundreds to thousands of kilometers (Le Pichon *et al.*, 2002; Liszka, 1978); however, in a previous study, Cates and Sturtevant (2002) showed that ray-tracing through a one-dimensional profile of wind and sound speeds underpredicted the extent of the sonic boom coverage.

The present study focuses on regions of sound exposure to Atlantis' supersonic descent and arrival times for sensors in these areas. We compare observations of infrasound arrivals at seismic and infrasound stations covering a region extending from 31° to 40°N and from −112° to −124°W with predictions based on ray-tracing through a detailed three-dimensional (3D) atmospheric model. Our aim is to evaluate the success of atmospheric models and ray-based infrasound propagation modeling in predicting the observations.

The atmospheric specifications, seismic network availability, and shuttle trajectory data for this study are discussed in Sec. II. A model of sound excitation by a supersonic object, as well as the subsequent transmission of infrasonic waves through a realistic windy atmosphere, are described in Sec. III. This is followed by a description of seismic and infrasound recordings of the Atlantis re-entry. The model predictions are compared to the observations in Sec. V, and we conclude with a discussion about the implications of our observations to the accuracy of the atmospheric models.

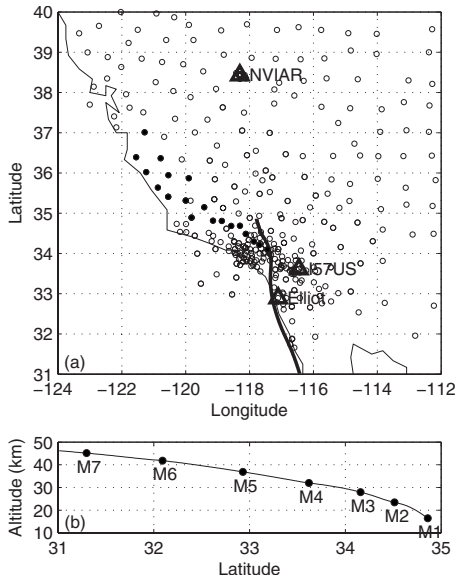


FIG. 1. (a) Map of the study area showing stations and the shuttle's path. Infrasound stations are represented by triangles and labeled. Seismic stations are represented by small circles. Wave forms of stations marked by filled circles are plotted in Fig. 8. The supersonic portion of the shuttle trajectory is shown by the dark solid line. (b) The shuttle altitude as a function of latitude is shown with associated Mach numbers.

II. DATA AND ATMOSPHERIC SPECIFICATIONS

This study combines a hybrid atmospheric specification at the time of the shuttle's re-entry with seismic and infrasound wave forms recorded on the ground and detailed telemetry on the shuttle itself.

A. Seismic data

Several seismic networks were operating in southern California, Nevada, and Mexico during the re-entry (Fig. 1). The most extensive network was the USArray (Meltzer *et al.*, 1999), which comprised over 300 stations and operated continuously through the re-entry. Deployment of the transportable stations within this array began in 2005 in southern California, and stations are scheduled to be removed in late 2007 for redeployment elsewhere. The June 22 re-entry of Atlantis is the only space shuttle landing that occurred in southern California during this time period. The ten-element NVAR seismic network in Nevada operated as part of USArray and is located 400 km north of the point at which the shuttle passed below Mach 1. NVAR has an aperture of 5 km. Several stations in northern Baja California were available as well as continuous and triggered stations operated by the California Institute of Technology (Caltech, Pasadena, CA). Ground velocity was recorded at between 20 and 100 Hz on three components by most stations in these networks. In total, there were about 350 stations operating within the study area.

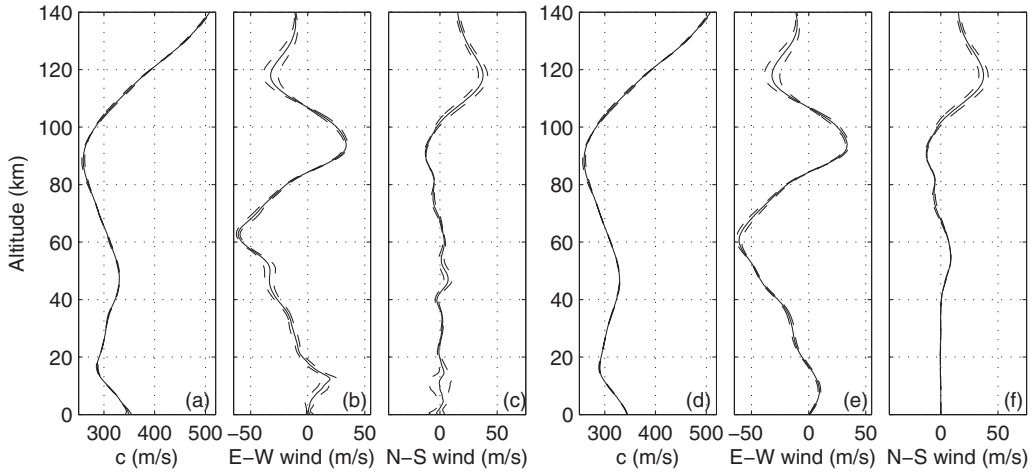


FIG. 2. (a) The mean sound speed as a function of altitude over the study area is shown by the solid line for the G2S model. The variability in the sound speeds over the study region is shown by the dashed lines, which indicate +/− one standard deviation in sound speeds at each altitude. (b) The mean zonal wind speed profile (positive eastward) over the same region is shown by the solid line, with the variability in values over the study region shown by the dashed lines. (c) Same as for (b) but for the mean meridional wind speed profile (positive northward). (d)–(f) show the corresponding values based on climatology data only.

B. Infrasound data

Infrasound stations located at Camp Elliott (collocated with seismic station 109C just northeast of San Diego), Anza-Borrego desert (I57US), and Nevada (NVIAR) recorded infrasound from the re-entry [Fig. 1(a)]. NVIAR was the only station to record the signals on enough array elements to determine the direction to the source. NVIAR is also colocated with the inner elements of the NVAR network and has an aperture of 1.6 km. The infrasound instrumentation at Camp Elliott [optical fiber infrasound sensor (OFIS)] is unique in that it uses fiber optic laser interferometry to resolve minute changes in pressure while reducing incoherent wind noise and turbulence by averaging pressure along a line (Zumberge *et al.*, 2003). The sensor has a flat instrument response from 0.01 to 100 kHz, so it is ideally suited to record an *N*-wave. The OFISs were not arranged in an array configuration at the time of the re-entry.

C. Shuttle telemetry

Shuttle trajectory data consist of a series of coordinated universal time (UTC) times and associated coordinates in the Earth centered Earth fixed reference system reported at 1 s intervals from over the Indian Ocean to the landing at Edwards AFB. These coordinates were converted to latitude, longitude, and altitude in the WGS-84 reference frame. The speed of the shuttle decreased from approximately Mach 6 near the United States-Mexico border to subsonic just south of Edwards AFB. The altitude decreased from over 40 km at Mach 6 to 16 km at Mach 1 [Fig. 1(b)]. There is a small error in the position reported by the GPS receiver located in the shuttle cockpit due to plasma effects. The 67% location uncertainty over the study area ranged from less than 25 m to less than 50 m.

D. Atmospheric specifications

The Naval Research Laboratory (NRL) ground to space (G2S) model of Drob *et al.* (2003) was used to provide background sound speeds and wind fields from 31°N to 40°N and from −124°W to −112°W, for 18:00 GMT, June 22, 2007. The G2S data processing system combines available global operational atmospheric numerical weather data from the National Oceanic and Atmospheric Administration Global Forecast System (NOAA-GFS) and the National Aeronautics and Space Administration Goddard Earth Observing System (Versions 4) (NASA-GEOS4) system (e.g., Kanamitsu, 1989; Kalnay *et al.*, 1990; Bloom *et al.*, 2005) with the NRL mass spectrometer, incoherent scatter model, 2000 version, horizontal wind model, 1993 version (NRL MSISE-00/HWM-93), and upper atmospheric empirical models (Picone *et al.*, 2002, Hedin *et al.*, 1996). The G2S model merges these data via a vector spherical harmonic transform and function fitting methodology between altitudes from the surface to approximately 250 km. The horizontal information content of the hybrid specifications varies from $1° \times 1°$ resolution in the lower atmosphere to $5° \times 5°$ resolution in the upper atmosphere. In the vertical direction the resolution varies from 350 m to a few kilometers in the upper atmosphere. Single point inverse vector spherical harmonic transforms combine with cubic spline interpolation to synthesize arbitrarily defined 3D data cubes of the winds and adiabatic sound speed fields in altitude coordinates at any desired output resolution.

The vertical sound and wind speed profiles extend up to an altitude of 140 km from the resulting G2S hybrid specification plus the corresponding NRL MSISE-00/HWM-93 climatological models (Fig. 2). To first order the primary aspects of the infrasound propagation characteristics are governed by the local vertical gradients of atmospheric fields (Drob *et al.*, 2003). The variability in sound and wind speeds over the study region for each model is indicated by the dashed curves. The ambient horizontal wind and sound speed gradients have a second order influence on the propagation; however, when integrated over long distances (greater than 500 km), significant changes in the local vertical structure cannot always be ignored (Drob *et al.*, 2003).

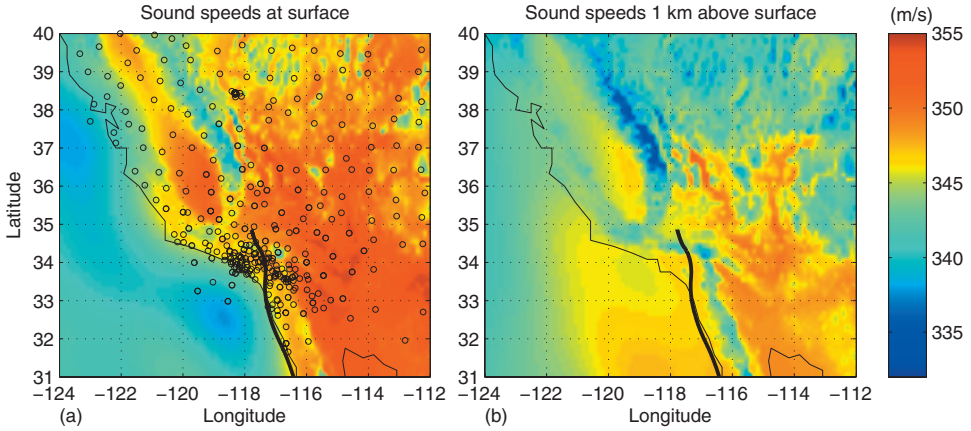


FIG. 3. (a) Sound speeds from G2S atmospheric data on the surface and (b) at a height of 1 km above the ground. Also shown are station locations and the shuttle trajectory down to Mach 1. NVAR is the dense collection of stations near 38°N and 118°W. The increase in sound speed with altitude over the ocean indicates the presence of a marine inversion layer.

Atmospheric specifications based on both models indicate that during the time of the event the adiabatic sound speeds at low altitudes are greater than those within the stratosphere. Thus adiabatic sound speed gradients alone are insufficient to cause sound to be refracted to the ground. At midlatitudes, the efficiency of stratospheric ducting is strongly influenced by seasonally varying zonal winds. During the time of this event, the zonal winds are reasonably strong to the west near the stratopause, approaching 60 m/s at an altitude of 60 km. Thus the effective sound speed (the adiabatic sound speed plus the wind speed in the direction of propagation) is increased for propagation to the west but decreases for propagation to the east. Therefore stratospheric ducting is predicted only for propagation to the west. Acoustic refractions from the large sound speed gradients in the thermosphere are also predicted. However, they are typically observed only for large events or for sensitive infrasound sensors (e.g., Balachandran *et al.*, 1977) since acoustic absorption increases with altitude and frequency (Sutherland and Bass, 2004).

As mentioned, in some instances range dependence can also be important in determining the infrasound propagation characteristics from source to receiver. A more detailed view of the G2S atmospheric model for this date and time is shown in Fig. 3 for sound speeds at the surface and at a height of 1 km. The marine inversion layer over the Pacific Ocean does not exist over the land mass, and infrasound may scatter into or out of this duct at the boundary.

III. SOURCE AND INFRASOUND PROPAGATION MODELS

The motion of a solid object through the atmosphere displaces air, causing a pressure disturbance that travels at the local sound speed. Pressure waves interfere constructively near any supersonic object, which creates a shock wave with a shape that depends on the object's geometry and speed (Whitham, 1974). Nonlinear effects near the shuttle cause the acoustic waves to steepen, which results in the characteristic *N*-wave generated by supersonic aircraft. The pressure disturbance propagates at right angles to the shock wave front near the aircraft. At greater distances, the pressure disturbance propagates acoustically with sound transmission governed by winds and sound speed gradients in the atmosphere.

Our model relies on the assumption that the shock wave governs only the allowable initial propagation angles and that the subsequent propagation of acoustic energy is linear from the shuttle trajectory to receivers on the ground. Linear propagation implies that the magnitude of the pressure perturbation due to the passage of the sound wave is very small compared to the ambient pressure. Under this approximation, the sound speed is independent of the pressure amplitude, and standard acoustic propagation methods may be used to model the sound transmission; here, we use ray-tracing methods. For Atlantis, which has a slender conically shaped nose, we make the standard assumption that the shock wave front may be represented as a cone with a half-angle of $\alpha = \sin^{-1}(c/v)$ where c is the sound speed and v is the shuttle velocity. The shuttle trajectory and the Mach cone angle α thus define the initial propagation direction for the acoustic wave front at any given point along the shuttle's trajectory. We assume that the shape of the cone, as well as the initial propagation direction, is not significantly affected by the winds along the shuttle's path or the shape of the shuttle's nose. A similar approach was used by Cates and Sturtevant (2002).

We approximate the source as a distribution of Mach cones at points along the shuttle trajectory, as in (Le Pichon *et al.*, 2002). Each source point is offset by location and time as determined from the telemetry data. For a coordinate system in which Atlantis travels horizontally and directly northward, allowable grazing angles, g, for rays perpendicular to the Mach cone are defined by

$$g = \sin^{-1}(\cos\alpha \cos\beta),$$

where g is defined as positive upward with respect to the horizontal axis and β is the interior angle within the Mach cone, measured clockwise from vertical as shown in Fig. 4. Feasible grazing angles for rays perpendicular to the Mach cone range from $-\cos^{-1}(c/v)$ to $+\cos^{-1}(c/v)$ at any given point along the shuttle's path. The corresponding ray azimuth with respect to north is given by

$$\theta = \sin^{-1}(\cos\alpha \sin\beta/\cos g).$$

The azimuths and grazing angles for a more general shuttle trajectory are derived by 3D coordinate rotation using the shuttle's azimuth and angle of descent at each point.

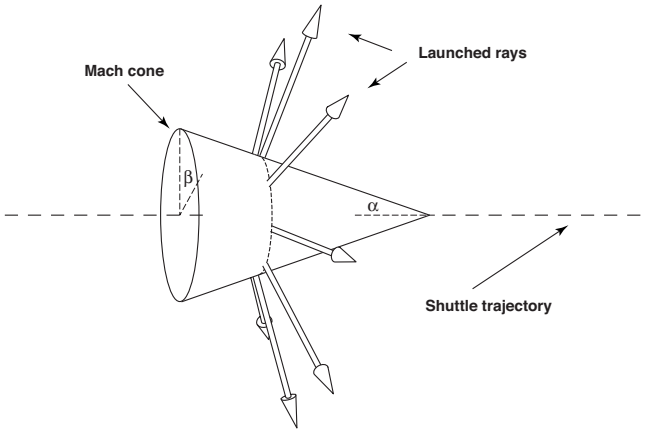


FIG. 4. Rays are launched at right angles to the Mach cone. The Mach cone angle α and interior angle β are indicated.

Each ray is launched at a right angle to the Mach cone (Fig. 4), at a position given by the shuttle's GPS receiver, and is propagated through the G2S model according to acoustic refraction laws in a windy atmosphere; these laws obey Fermat's principle of least time in an advected medium (Thompson, 1972; Ostashev *et al.*, 2001). Our ray method allows for both vertical and lateral variations in adiabatic sound speeds and winds, as well as reflection points along Earth's surface. Rays are launched at even increments in β, the interior cone angle.

Longitudinal propagation through the G2S model is illustrated in Fig. 5 for a single point on the shuttle's trajectory, not taking into account cone geometry. For illustration, we make the assumption that the ground surface along this profile is flat and at sea level elevation. We also have not considered rays that turn in the thermosphere above 120 km. Points west of 119°W lie over the Pacific Ocean. Rays are initiated within the vertical plane over a series of angles relative to the horizontal ranging from directly down to directly up. Black lines illustrate rays with end points on the ground.

As shown in Fig. 5, steeply downward-going rays from the shuttle form a primary sound carpet across both directions. There is no ducting of infrasound energy between the ground and upper atmosphere for propagation to the east due to the presence of the westward summer-time stratospheric jet. For westward propagation, rays launched at shallow upward angles refract from the stratosphere back to the ground to form a secondary sound carpet. Multiple refractions and reflections between the stratosphere and the ground, and within the lowest 1 km of the troposphere, carry the sound farther from the source. The lower panel of Fig. 5 illustrates the reduced travel times corresponding to these rays. Rays are launched at uniform intervals; thus, the G2S atmospheric specifications for this event predict far fewer rays, and hence lower infrasonic levels, between the primary carpet and the first westward propagating stratospheric arrivals. Farther from the source, some of the energy originally ducted between the ground and stratosphere becomes trapped in a marine inversion layer over the Pacific Ocean. This inversion layer has the effect of increasing infrasound levels at the surface. The lower panel of Fig. 5 illustrates multiple arrivals for points west of 120.2°W. The separate arrivals imply that acoustic phase velocities associated with stratospheric ducting arrive later than for tropospheric ducting. Consequently, later arrivals are associated with acoustic energy that has been stratospherically ducted over a greater portion of the travel path.

Multipathing and acoustic shadow zones are further investigated in Fig. 6. Results from atmospheric specifications based on the G2S model and the standard climatology are shown. End points at which the rays reached the sea level reference surface are marked by dots. Rays were launched at equal increments of 0.2° over the interior of the Mach cone and at 6 s intervals along the shuttle path, so regions where the density of points is low between direct and stratospheric arrivals suggest low infrasound levels. Acoustic shadow zones exist beyond the primary sound carpets to the east of the shuttle trajectory and to the NW of the landing point. As shown, the ray propagation method predicts a sharp edge to the primary sound carpet east of the shuttle transit path for both sets of atmospheric specifications. The transition is less pronounced to the west due to stratospheric and tropospheric ductings. The G2S model predicts that tropospheric ducting

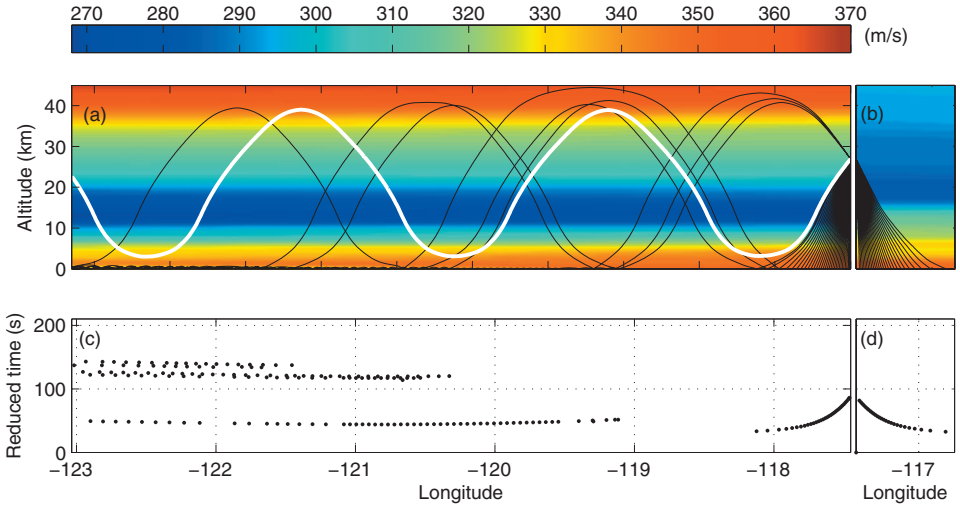


FIG. 5. (a) Ray paths from 34.25°N and 118.5°W for propagation to the west. The white line indicates one of the many ray paths that do not reach the ground. Background colors indicate the effective sound speed, given by the sum of the adiabatic sound speed and the wind velocity in the direction of propagation. Note that these are equal to sound speeds in an advected medium only for horizontal propagation. (b) Rays for propagation to the east. Identical color scales are used in (a) and (b). (c) Reduced travel times for propagation to the west and (d) east. Reduced travel times are given by $(t-|r|/v_r)$ as a function of longitude, where t is the travel time in seconds along the ray path, $|r|$ is the range from the source, and v_r is the reducing velocity, set to v_r=335 m/s here.

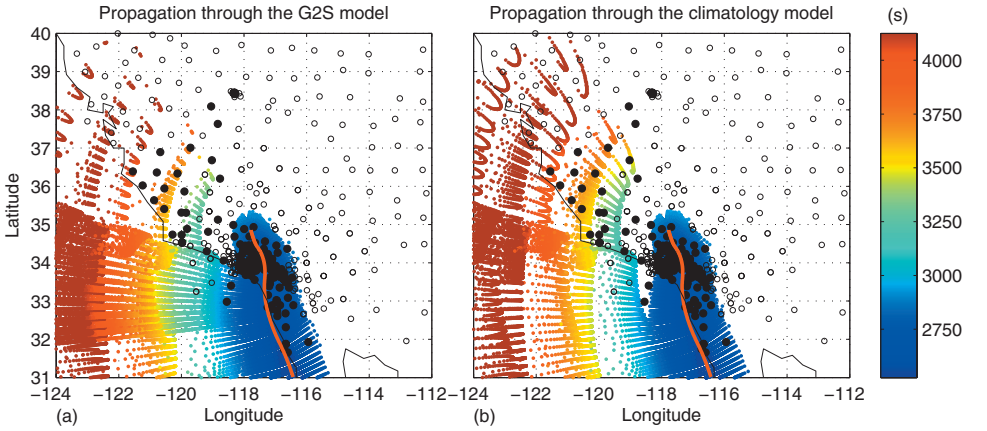


FIG. 6. Maps of ray end points that reach the ground, for rays starting along the supersonic portion of the shuttle trajectory, are shown in red. We have ignored rays that turn in the thermosphere above 120 km; these would arrive significantly later for stations located over 250 km from the shuttle path. The rays were propagated through atmospheric specifications provided by the (a) G2S model and by (b) climatology data. The end points are color-coded according to the predicted arrival time in seconds after 1900 GMT. For comparison, the shuttle speed drops below Mach 1 at 2732 s after 1900 GMT. Small circles indicate station locations that did not register detections; filled circles indicate stations that recorded an arrival. Rays were launched at even intervals, so regions where ray densities are low indicate either low infrasound levels (e.g., regions between stratospheric arrivals) or acoustic shadow zones (beyond the primary carpets to the east and north).

is significant only over the ocean; the climatology predicts only stratospheric returns. Shadow zones are predicted farther to the north-NW for the atmospheric model based on climatology. Note that these differences arise from relatively small differences in the wind profiles at altitudes of less than 60 km, as indicated in Fig. 2.

IV. OBSERVATIONS

The observations for this paper are based on continuous and triggered recordings made at over 130 seismic stations distributed across the study area as well as continuous recordings made at 3 infrasound stations. We have divided our recordings into two classes. Near-field recordings were made by infrasound and seismic stations located beneath the shuttle or immediately to the side (within about 100 km of the shuttle on its closest pass). Acoustic energy propagated directly downward from the shuttle to these stations. Therefore signals at these stations can be unambiguously associated with the shuttle. At greater distances, it was more difficult to associate acoustic signals at seismic stations with the shuttle's passage as station density in the outlying areas was lower.

A. Near-field recordings

The shuttle was recorded by 109 seismic stations and 2 infrasound stations located in a 200 km wide corridor beneath Atlantis north of the United States-Mexico border. Many of the seismic recordings show a remarkable degree of similarity even though they recorded the shuttle at speeds ranging from over Mach 5 down to Mach 1 at distances ranging from 0 to over 80 km east and west of the shuttle's track [Fig. 1(a)]. Associating acoustic arrivals from the shuttle at these stations was straightforward for a couple of reasons. First, most signals were recorded with a high signal-to-noise ratio (SNR). Second, the station density was high beneath the shuttle from the United States-Mexico border to the point at which the shuttle dropped below Mach 1 near Edwards AFB, permitting us to observe the acoustic phase moving out between stations.

The acoustic signal near the shuttle is an *N*-wave; the equivalent seismic velocity response is a doublet. The seismic station 109C was colocated with OFISs at Camp Elliott near San Diego. Figure 7 shows the OFIS acoustic and adjacent seismic recordings at this location. The *N*-wave is recorded well by the broadband OFIS. The collocated broadband seismometer (STS-2) recorded a modified version of the *N*-wave. Also seen in the recorded seismic velocity is a Rayleigh wave that precedes the acoustic signal at this location. Rayleigh waves have been attributed to acoustic-to-

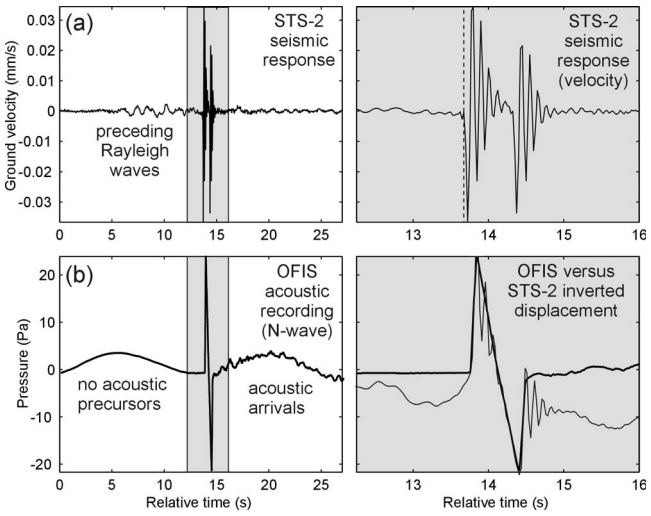


FIG. 7. Comparison between the (a) seismic and (b) infrasound recordings for US Array station 109C (Camp Elliott). The vertical dashed line marks the travel time pick based on the seismic data. The integral of the seismic velocity record is plotted (as the gray line) with the infrasound recording in the lower right panel.

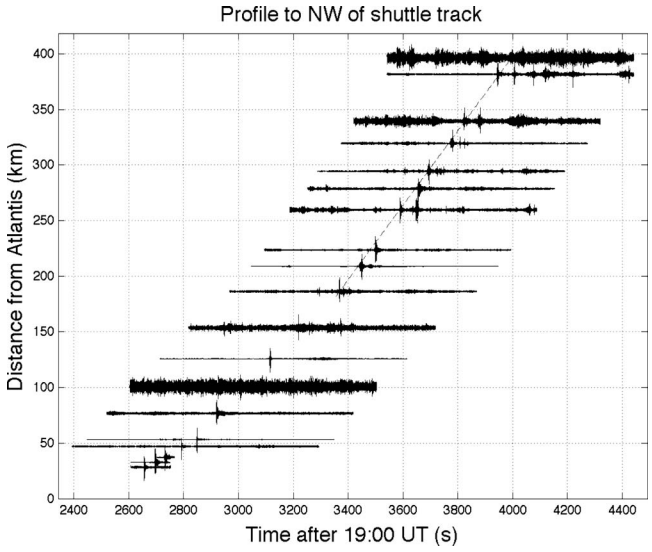


FIG. 8. Bandpassed recordings (2–8 Hz) along a profile to the NW of the shuttle, indicated by filled circles in Fig. 1. The wave forms are plotted as a function of distance from the nearest point on the shuttle path, which was located just south of Edwards AFB where the shuttle slowed to Mach 1. The increasing complexity of recordings past 225 km may indicate multipathing. The dashed line delineates a propagation velocity of 335 m/s.

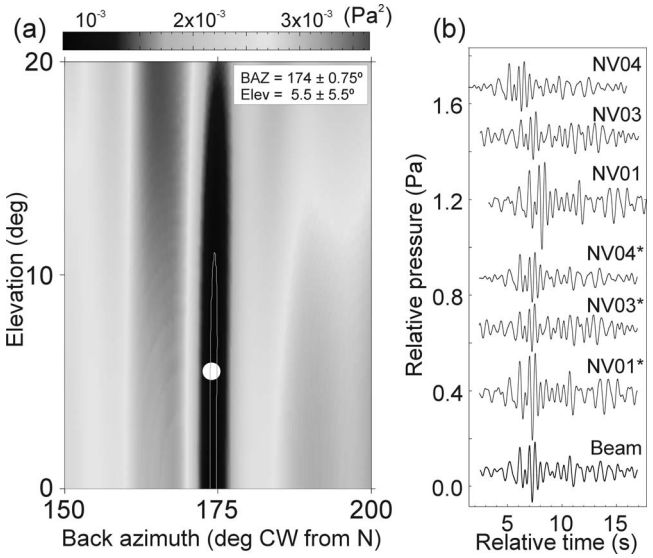


FIG. 9. Time domain beamforming for infrasound station NVIAR, Nevada. A 0.8–3.0 Hz bandpass filter is applied to the data. The array aperture is 1600 m, which can resolve subhorizontal signals down to about 0.1 Hz. (a) The 95% confidence region is outlined by a gray line. (b) The time-shifted wave forms.

seismic coupling near the shuttle (e.g., Kanamori *et al.*, 1991) and are easily distinguished from the infrasound arrival due to their lower frequencies and amplitudes.

B. Distant recordings

Wave form recordings of the shuttle made more than 100 km from the shuttle's path are more structured. Due to longer propagation paths, variable noise levels at the recording stations and the greater spacing between stations, these signals are poorly correlated. Wave forms at these stations are longer in duration and, in some instances, appear as a series of multipathed arrivals. Similar behavior for the seismic observations of the Washington state bolide event was observed (Arrowsmith *et al.*, 2007). To isolate signals from the shuttle at these stations, we plotted recordings as a function of distance of each station from the closest pass of the shuttle. In one example of such a record section, shown in Fig. 8, there is a distinct onset of energy at the distant stations located to the NW of the shuttle's track moving out at a speed of approximately 335 m/s (dashed line). In each case, we picked travel times based on the first onset of the impulsive energy (i.e., after the Rayleigh wave).

Seismic networks can be useful for detecting high amplitude infrasound signals but weaker infrasound signals, or those that propagate subhorizontally, may not generate recognizable seismic phases. For example, the NVIAR infrasound array is located at the edge of the seismic signal detection zone in Nevada and recorded signals (with an average SNR of 16 dB) that were not coherent across the much larger collocated NVAR seismic array.

Although three infrasound signals can be observed on most of the infrasound array, the signals may be obscured by intermittent wind noise. The second signal has the highest SNR. Time domain beamforming is used to estimate the second infrasound signal's back azimuth and elevation angle (Fig. 9). The acoustic wave speed is calculated from the measured air temperature assuming that the winds are negligible. A grid search is applied over trial back azimuth and elevation angle to minimize the sum of squares of the misfit between time-shifted wave forms, permitting a rigorous search of the solution space for local minima. The 95% confidence contour is derived assuming the sum of squares reflects a chi-squared noise process and using the technique of Silver and Chan (1991) to estimate the average number of degrees of freedom for the filtered wave forms. The back azimuth and elevation angle are estimated to be $174 \pm 0.75°$ and $5.5 \pm 5.5°$. This suggests an apparent speed across the array of 351 m/s. At such a low elevation angle, an approximately 10 m/s difference in north-south wind speed is required to explain a 10° inaccuracy in elevation angle. Three incoherent seismic signals are observed across the NVAR array, with energy packets that align along the 351 m/s moveout line (Fig. 10). The outlying elements of NVAR are not shown due to high noise (NV11, NV31) or poor data quality (NV32, NV33). Projection of the back azimuth uncertainty envelope from the middle of the infrasound array defines an 11 km wide swath that blankets Edwards AFB and the Mach 1.0 part of the trajectory. This may suggest that deceleration and consequential flattening of the Mach cone are the sources of the infrasound signals at this location in Nevada.

The detections and arrival times at all stations are summarized in Fig. 11. The accuracy in the arrival times for the near-field stations is less than 0.1 s due to the high SNR, sharp signal onsets, high coherence of the doublet signals between the closely spaced stations, and the acoustic moveout. At the distant stations the signals were highly variable and less defined, resulting in arrival time errors on the order of 0.5–1 s. Stations not highlighted in Fig. 11 did not record a detectable signal.

Many stations beneath the shuttle's track recorded a signal (Fig. 11). To the east, there is a sharp boundary between

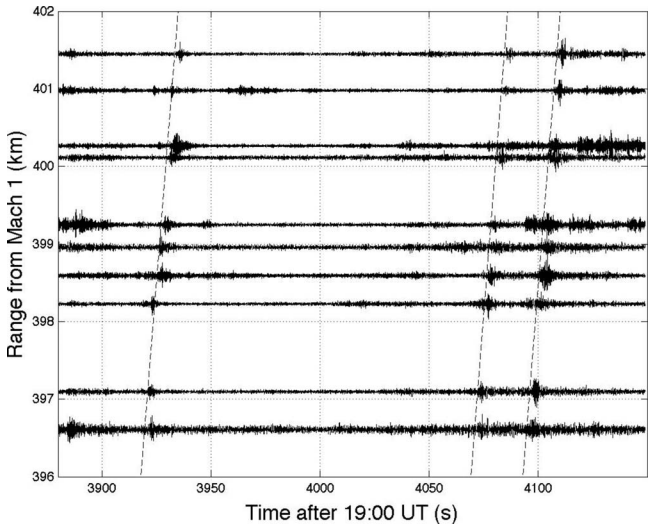


FIG. 10. NVAR seismic signals plotted with respect to time and distance to the point at which the shuttle slowed to Mach 1. The dashed lines have a slope of 352 m/s. These traces were bandpass filtered from 0.5 to 5.0 Hz.

the primary sound carpet and the shadow zone, with no signals observed farther to the east. Stations on two of the five islands to the west recorded signals. Surprisingly, most stations less than 400 km to the NW of the shuttle recorded the event. As discussed earlier, the colocated NVAR seismic network and NVIAR infrasound array approximately 400 km north of Edwards AFB clearly recorded the shuttle. The lack of detections at stations to the NW suggest the presence of regions of low ensonification or a shadow zone although it may be that some stations did not record a signal due to high noise or poor acoustic-to-seismic coupling. Our ability to map out the precise boundary of the shadow zone is also limited by the station density. Finally, it should be noted that the observed arrival times are inconsistent with thermospheric arrivals, which are predicted to arrive later than the stratospheric arrivals. Thermospheric arrivals were not observed in the frequency bands considered by this study.

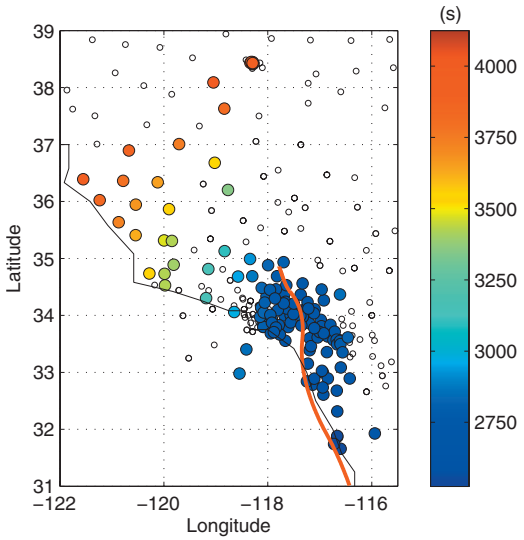


FIG. 11. Map of stations at which infrasound arrivals were detected. The stations are color-coded to show the time of the first arrival in seconds after 1900 GMT. Small circles show where signals were not detected. Compare to Fig. 6, which shows predicted travel times at the same color scale.

V. COMPARISON BETWEEN OBSERVATIONS AND PREDICTIONS

Comparison of Fig. 11 with Fig. 6 indicates qualitative agreement between predictions and observations. First, we predict a sharp edge to the primary acoustic carpet to the east of the shuttle with no propagation after that, in agreement with observations. The atmospheric specifications based on the G2S model predict tropospheric ducting and downward refraction over the ocean to the west, which agrees with the signals observed at the island stations. Ray modeling based on both sets of atmospheric specifications predicts alternating areas of high and low ensonifications over land to the NW, with most stations not in a position to record detectable signals; this conflicts with the detections over a broad region to the NW. We also accurately predict a total absence of signals to the north of Edwards AFB and beyond the primary carpet to the east of the shuttle trajectory. Finally, the 335 m/s moveout velocity, shown in Fig. 8 for arrivals to the NW of the shuttle track, shows close agreement with the predicted moveout speed for arrivals to the west (Fig. 5).

Travel times from the shuttle trajectory to each site were computed in order to make a quantitative comparison between predications and observations. Rays were launched at a greater ray density than illustrated in Fig. 6 to match observation points with predicted travel time. Families of rays with end points within a given radial cutoff distance d_c from the station and with similar travel paths were binned to yield travel times. The cutoff distance d_c was set either to 6 km, or the distance subtended by a 2° angle at a range of $r_{\min}$, whichever is larger, where $r_{\min}$ is the shortest range between the station and shuttle trajectory. We assume that travel times vary with radial distance from source to receiver, and only negligibly with azimuth, over the distance d_c. For each station, we therefore corrected the travel times using a reducing velocity derived from a linear fit to a travel time versus overshoot curve.

Figure 12 shows the time residual between the predicted and observed arrival times, defined here as the predicted minus the observed arrival time. These residuals were computed both for atmospheric specifications. In the context of this discussion it should be noted that there are instances where the actual state of the atmosphere will be close to that of the climatological average, whereas at other times there can be significant departures; for example, during the time and location of the Buncefield oil depot explosion (Evers and Haak, 2007) where stratospheric wind velocities were 120 m/s. As shown in Fig. 2 for this event, the state of the lower and middle atmosphere is close to that of the climatological state. Both models indicate large negative residuals for both sets of atmospheric specifications at a longitude of about 120°W and at latitudes south of 36°N. Predicted arrival times within this region are too early by approximately 80 s. Comparison with the travel time predictions shown in Fig. 6 indicates that, for both atmospheric models, these points lie in a region where ray densities are comparatively low. At the station farthest to the NW (the second trace from the top in Fig. 8), the predicted time lagged the observed time by

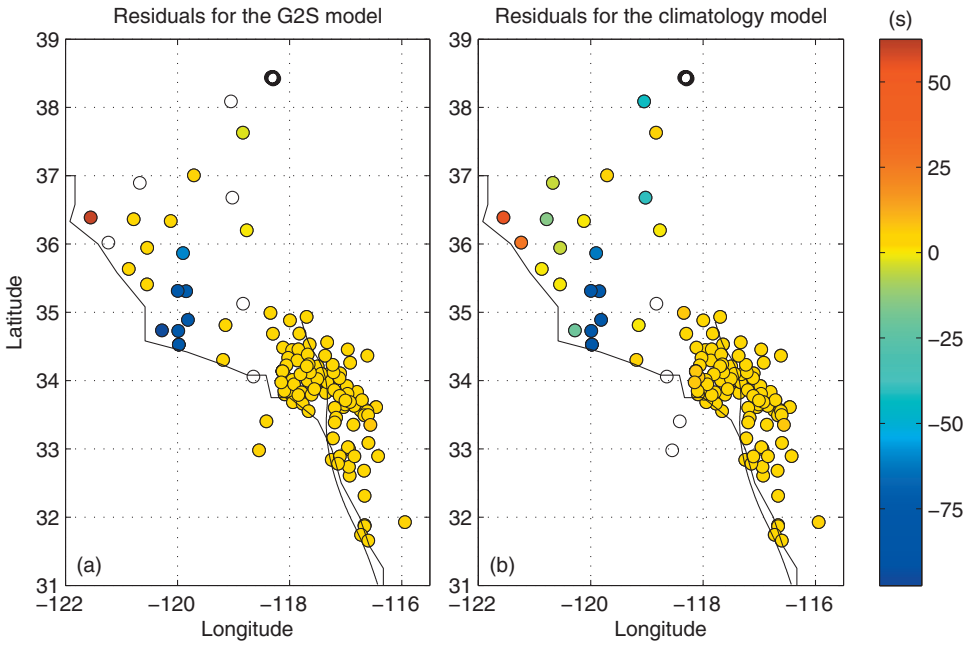


FIG. 12. Travel time residuals, in seconds, for atmospheric specifications based on the (a) G2S model and (b) climatology data. Positive residual values indicate that predicted arrival times are too late; negative residuals indicate that predicted arrival times are earlier than observed. Empty circles indicate stations at which observations were made but have no associated predictions for the given ray density used for computation of travel time predictions.

51–58 s for both atmospheric models. The predicted arrival times are consistent with the second arrival at that station, as shown in Fig. 8. Contrary to observations, both models predict that the Nevada stations are not far enough to the west to observe infrasound arrivals.

A significant feature of these results is the excellent agreement between predicted and observed arrival times for stations in regions where ray densities are predicted to be high. On average, observed arrival times were early by 2.6 s in these regions.

VI. DISCUSSIONS AND CONCLUSIONS

The supersonic descent of Atlantis above an area occupied by the USArray and other seismic networks on June 22, 2007 yielded an opportunity to evaluate atmospheric propagation modeling capabilities over a wide spatial, although not temporal region. The seismic networks across northern Mexico, California, and Nevada represented a 100- fold increase in station density over the available infrasound site density. The greater station availability for stations outside the primary sound carpet averted the usual difficulty of associating a seismic or infrasound signal with a particular event. For this reason, we were able to associate infrasound signals recorded on seismic sensors with Atlantis to distances of 400 km from the nearest point on the shuttle trajectory.

Comparison of predicted versus observed travel times shows agreement over much of the study area, as discussed in Sec. V. For the G2S model, travel time residuals are small within regions that are highly ensonified, as predicted by ray theory. This implies that the source model, infrasound propagation model, and G2S atmospheric specifications are adequate for predicting arrival times within these regions. It should be noted, however, that the agreement between observed and predicted arrival times at distances beyond a ground reflection point may have been somewhat fortuitous as we ignored terrain effects. In reality, scattering or deflection may occur at the reflection point if the ground surface is rough.

The results also suggest that the use of geometrical acoustics is inaccurate in computing arrival times where very low ray densities are predicted and, clearly, also within shadow zones. This is supported by the fact that ray theory relies on a high frequency approximation to wave propagation, which is valid when sound speeds vary on a scale length much larger than the propagation wavelength. This approximation starts to break down at infrasonic frequencies for typical atmospheric conditions. A more complete description of infrasound propagation includes finite wavelength effects such as diffraction, scattering (Embleton, 1996), and surface waves (Attenborough, 2002). These effects lead to much greater penetration of acoustic energy into shadow zones at amplitudes than predicted by ray theory (Embleton, 1996; Attenborough, 2002). This argument is further supported by the fact that infrasound detections have also been reported at distances of up to tens of kilometers within areas predicted by ray theory to be shadow zones (e.g., Ottemöller and Evers, 2008).

The small positive bias in the arrival time residuals within the shuttle's primary sound carpet indicates either a source effect or a bias in the adiabatic sound speeds. Potential errors in the source model point to a slight overestimate of the arrival times. First, there may be an error in the shape of the Mach cone used in our modeling. Equation (9.74) of (Whitham, 1974) indicates that at high Mach numbers, the true Mach cone angle is slightly greater than the simplifying assumption of $\alpha = \sin^{-1}(c/v)$ used here. This would lead to a different take-off angle from the Mach cone. Second, infrasound propagation near the shuttle is nonlinear as the result of increased sound speeds from the local overpressure. However, these effects on the order of several tenths of a second are not enough to account for the average 2.6 s bias observed. The bias in the adiabatic sound speeds within the atmosphere would have to be on the order of 5 m/s to yield the observed arrival times.

Finally, note that relatively small variations in the wind speed profiles can yield significant variations in predicted infrasound propagation. As shown in Fig. 2, the climatological wind speed profiles are smoother than those based on the G2S model. This yields differences in the pattern of ensonification at the ground level as shown in Fig. 6; highly ensonified regions are broader for the climatological model and extend along a line farther to the north-NW than for the G2S model. Several stations lie in the predicted acoustic shadow zone just beyond this line, including the Nevada stations where stratospheric arrivals were observed by both sensors. Instantaneous uncertainties of the G2S specification in the stratosphere are on the order of 10–15 m/s. This uncertainty may be sufficient to account for the lack of arrival time predictions for stations to the north-NW as well as the mismatch of the predicted surface skip locations of the stratospheric rays. More generally, insonification of the shadow zone could be due to scattering or refraction by small scale atmospheric inhomogeneities. The inhomogeneities have scale lengths comparable to infrasonic wavelengths and have been shown to affect infrasound propagation to large distances from a source (Kulichkov *et al.*, 2004).

ACKNOWLEDGMENTS

We thank J. F. Zumberge of the Jet Propulsion Laboratory and S. V. Murray of the Johnson Space Center for the shuttle time and position data. The NVIAR (infrasound) data were provided by the U.S. Army Space and Missile Defense Command. Support for the NVIAR station comes from the Southern Methodist University and the National Center for Physical Acoustics at the University of Mississippi. We are grateful to Raul Castro Escamilla at CICESE for providing data from the Mexican seismic stations. This study benefited greatly from the Data Management System offered by the Inc. Research Institutions for Seismology. We would also like to thank the NASA Goddard Space Flight Center, Global Model and Assimilation Office, and the NOAA National Centers for Environmental Predication for making their numerical weather prediction (NWP) products available for this scientific research. The High-Performance Wireless Research and Education Network provided wireless network access to the infrasound sensors at PFO and Camp Elliott. We would like to thank our three anonymous reviewers for their constructive feedback. This work was funded by the U.S. Army Space and Missile Defense Command under the University of Mississippi Subcontract No. 07-08-013. Auxiliary support is also provided to NRL by the Office of Naval Research. We dedicate this work to Hank Bass.

Arrowsmith, S. J., Drob, D. P., Hedlin, M. A. H., and Edwards, W. (**2007**). "A joint seismic and acoustic study of the Washington State bolide: Observations and modeling," J. Geophys. Res. **112**, D09304.

Attenborough, K. (**2002**). "Sound propagation close to the ground," Annu. Rev. Fluid Mech. **34**, 51–82.

Balachandran, N., Donn, W., and Rind, D. (**1977**). "Concorde sonic booms as an atmospheric probe," Science **197**, 47–49.0036-8075

Bloom, S., da Silva, A., Dee, D., Bosilovich, M., Chern, J.-D., Pawson, S., Schubert, S., Sienkiewicz, M., Stajner, I., Tan, W.-W., and Wu M.-L. (**2005**). "Documentation and validation of the goddard earth observing system (GEOS) data assimilation system—Version 4," NASA Report No. TM-2005-104606, Technical Report Series on Global Modeling and Data Assimilation, Greenbelt, MD.

Cates, J. E., and Sturtevant, B. (**2002**). "Seismic detection of sonic booms," J. Acoust. Soc. Am. **111**, 614–628.

Cochran, E. S., and Shearer, P. M. (**2006**). "Infrasound events detected with the southern California seismic network," Geophys. Res. Lett. **33**, L19803.

Drob, D. P., Picone, J. M., and Garces, M. (**2003**). "Global morphology of infrasound propagation," J. Geophys. Res. **108**, ACL13.1–ACL13.12.0148-0227

Embleton, T. F. W. (**1996**). "Tutorial on sound propagation outdoors," J. Acoust. Soc. Am. **100**, 31–48.

Evers, L. G., and Haak, H. W. (**2007**). "Infrasonic forerunners: Exceptionally fast acoustic phases," Geophys. Res. Lett. **34**, L10806.

Gibbons, S., Ringdal, F., and Kvaerna, T. (**2007**). "Joint seismic-infrasonic processing of recordings from repeating source of atmospheric explosions," J. Acoust. Soc. Am. **122**, EL158–EL164.

Hedin, A. E., Fleming, E. L., Manson, A. H., Scmidlin, F. J., Avery, S. K., Clark, R. R., Franke, S. J., Fraser, G. J., Tsunda, T., Vial, F., and Vincent, R. A. (**1996**). "Empirical wind model for the upper, middle, and lower atmosphere," J. Atmos. Terr. Phys. **58**, 1421–1447.0021-9169

Ishihara, Y., Furumoto, M., Sakai, S., and Tsukada, S. (**2004**). "The 2003 Kanto large bolide's trajectory determined from shockwaves recorded by a seismic network and images taken by a video camera," Geophys. Res. Lett. **31**, L14702.0094-8276

Kalnay, E., Kanamitsu, M., and Baker, W. E. (**1990**). "Global numerical weather prediction at the National Meteorological Center," Bull. Am. Meteorol. Soc. **71**, 1410–1428.

Kanamitsu, M. (**1989**). "Description of the NMC global data assimilation and forecast system," Weather Forecast. **4**, 335–342.0882-8156

Kanamori, H., Mori, J., Anderson, D. L., and Heaton, T. H. (**1991**). "Seismic excitation by the space shuttle Columbia," Nature (London) **349**, 781–782.0028-0836

Kanamori, H., Mori, J., Sturtevant, B., Anderson, D. L., and Heaton, T. H. (**1992**). "Seismic excitation by space shuttles," Shock Waves **2**, 89–96.

Kulichkov, S. N. (**2004**). "Long-range propagation and scattering of low-frequency sound pulses in the middle atmosphere," Meteorol. Atmos. Phys. **85**, 47–60.0177-7971

Langston, C. (**2004**). "Seismic ground motions from a bolide shock wave," Geophys. Res. Lett. **109**, B12309.0094-8276

Le Pichon, A., Garces, M., Blanc, E., Barthelemy, M., and Drob, D. P. (**2002**). "Acoustic propagation and atmosphere characteristics derived from infrasonic waves generated by the Concorde," J. Acoust. Soc. Am. **111**, 629–641.

Liszka, L. (**1978**). "Long-distance focusing of concorde sonic boom," J. Acoust. Soc. Am. **64**, 631–635.

Meltzer, A., Rudnick, R., Zeitler, P., Levander, A., Humphreys, G., Karlstrom, K., Ekström, G., Carlson, R., Dixon, T., Gurnis, M., Shearer, P., and van der Hilst, R. (**1999**). "USArray initiative," GSA Today **9**, 8–10.1052-5173

Millet, C., Robinet, J.-C., and Roblin, C. (**2007**). "On using computational aeroacoustics for long-range propagation of infrasounds in realistic atmospheres," Geophys. Res. Lett. **34**, L14814.0094-8276

Ostashev, V. E., Hohenwarter, D., Attenborough, K., Blanc-Benon, P., Juve, D., and Goedecke, G. H. (**2001**). "On the refraction law for a sound ray in a moving medium," Acta Acust. **87**, 303–306.

Ottemöller, L., and Evers, L. G. (**2008**). "Seismo-acoustic analysis of the buncefiled oil depot explosion in the UK, 2005 December 11," Geophys. J. Int. **172**, 1123–1134.

Picone, J. M., Hedin, A. E., Drob, D. P., and Aiken, A. C. (**2002**). "NRLMSIS-00 empirical model of the atmosphere: Statistical comparisons and scientific issues," J. Geophys. Res. **107**, 1468.

Silver, P. G., and Chan, W. W. (**1991**). "Shear-wave splitting and subcontinental mantle deformation," J. Geophys. Res. **96**, 16429–16454.0148-0227

Sutherland, L. C., and Bass, H. E. (**2004**). "Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **115**, 1012–1032.

Thompson, R. J. (**1972**). "Ray theory for an inhomogeneous moving medium," J. Acoust. Soc. Am. **51**, 1675–1682.

Whitham, G. B. (**1974**). *Linear and Nonlinear Waves* (Wiley, New York).

Zumberge, M. A., Berger, J., Hedlin, M. A. H., Hilt, R., Nooner, S., and Widmer-Schnidrig, R. (**2003**). "An optical fiber infrasound sensor: a new lower limit on atmospheric pressure noise between 1 Hz and 10 Hz," J. Acoust. Soc. Am. **113**, 2474–2479.

Acoustic mode coupling induced by shallow water nonlinear internal waves: Sensitivity to environmental conditions and space-time scales of internal waves

John A. Colosi
Department of Oceanography, Naval Postgraduate School Monterey, California 93943

(Received 1 October 2007; revised 6 March 2008; accepted 11 June 2008)

While many results have been intuited from numerical simulation studies, the precise connections between shallow-water acoustic variability and the space-time scales of nonlinear internal waves (NLIWs) as well as the background environmental conditions have not been clearly established analytically. Two-dimensional coupled mode propagation through NLIWs is examined using a perturbation series solution in which each order n is associated with nth-order multiple scattering. Importantly, the perturbation solution gives resonance conditions that pick out specific NLIW scales that cause coupling, and seabed attenuation is demonstrated to broaden these resonances, fundamentally changing the coupling behavior at low frequency. Sound-speed inhomogeneities caused by internal solitary waves (ISWs) are primarily considered and the dependence of mode coupling on ISW amplitude, range width, depth structure, location relative to the source, and packet characteristics are delineated as a function of acoustic frequency. In addition, it is seen that significant energy transfer to modes with initially low or zero energy involves at least a second order scattering process. Under moderate scattering conditions, comparisons of first order, single scattering theoretical predictions to direct numerical simulation demonstrate the accuracy of the approach for acoustic frequencies upto 400 Hz and for single as well as multiple ISW wave packets. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956471]

PACS number(s): 43.30.Bp, 43.30.Re, 43.60.Cg [WLS] Pages: 1452–1464

I. INTRODUCTION

Nonlinear internal waves (NLIWs) in shallow water environments have been shown numerous times to produce significant sound-speed perturbations which can have large impacts on acoustic transmissions (Zhou, Zhang, and Rogers, 1991; Headrick *et al.*, 2000a; Rouseff *et al.*, 2002; Chiu *et al.*, 2004; Duda *et al.*, 2004b; and Fredricks *et al.*, 2005). Together with the stochastic background of linear internal waves (Apel *et al.*, 1997), NLIWs impose the ultimate limitations to acoustic remote sensing in shallow water environments. Because in any geographic region, NLIW parameters vary significantly over time (Colosi *et al.*, 2001; Duda *et al.*, 2004a; Ramp *et al.*, 2004), it is important to understand the sensitivity of acoustic fields to this environmental variability. To this end a direct numerical sensitivity analysis could be undertaken (for example, see Frank *et al.*, 2004), or a theoretical analysis could be done. In this study we primarily choose the latter.

Direct numerical analysis of the impacts of NLIWs on shallow water acoustic propagation have been numerous, and often accompany observational results to aid in interpretation (Tielburger, Finette, and Wolf, 1997; Preisig and Duda, 1997; Duda and Preisig, 1999; Headrick *et al.*2000b; Rouseff *et al.*, 2002; Frank *et al.*, 2004; Chiu *et al.*, 2004; and Fredricks *et al.*, 2005). However, direct theoretical analyses of the NLIW problem have been very few in number. The seminal work of Zhou, Zhang, and Rogers (1991) used an oceanographically idealized sinusoidal train of NLIWs to successfully describe the enormous intensity fluctuations observed in the Yellow Sea. Also in seminal work, Preisig and Duda (1997) showed that appropriately chosen "square solitons," in which the sudden approximation was applied at the two abrupt interfaces, could be used to understand and predict coupling behavior in direct numerical simulations with oceanographically realistic internal solitary waves (ISWs).

However, to avoid these problems of approximating the NLIW structure, the approach taken here is to use the Dyson series from quantum mechanical perturbation theory (Dyson, 1949; Sakurai, 1985) to solve for the change in modal amplitudes caused by NLIWs, by only approximating the acoustic scattering physics. In the ocean acoustics context, the Dyson series solution was introduced by Dozier and Tappert (1978a)[1], and this method provides important physical insight into the coupling physics because the nth term in the series can be associated with nth order multiple scattering. For ease of analytic treatment the present analysis is confined to coupling in the depth-range space; azimuthal coupling is ignored. Hence, this model is intended to describe mode coupling when the acoustic path is nearly perpendicular to the NLIW front, though the precise angular region where azimuthal coupling can be ignored remains an important area of investigation (Badiey *et al.*, 2007, 2005; Frank *et al.*, 2004, 2005). Be that as it may, this two-dimensional (2D) approach provides powerful insight into the dependence of acoustic mode coupling on environmental conditions, like the profiles of sound speed, density, and attenuation, and the space-time scales of NLIWs. Thus, this paper establishes the rigorous theoretical foundations and clarifies several issues from the seminal work of Zhou, Zhang, and Rogers (1991),

Preisig and Duda (1997), and Duda and Preisig (1999), and presents several new results.

The key new results of this paper can be summarized as follows. The Dyson series terms reveal important resonance conditions that select specific NLIW horizontal wavenumbers contributing to the coupling of acoustic modes. Modal attenuation fundamentally alters the coupled mode physics, by broadening the resonances in wavenumber, thereby allowing more NLIW wavenumber components to interact with the modes. Significant and observable broadening will only occur at low frequencies, which are estimated to be less than a few hundred hertz (Hz). While Dozier (1983) and Creamer (1996) have addressed the effects of attenuation on mode interaction with a stochastic spectrum of internal waves, the fundamental broadening of the resonance conditions was not identified in their work.

Examination of the resonance conditions shows the sensitivity of mode coupling to the spatial scales and amplitude of NLIWs. At first order (single scattering) the interaction of modes n and m has its strongest contribution (resonance), where the NLIW wavenumber matches the beat wavenumber of the two modes (i.e. $k_{mn}=k_n-k_m$), and the strength of coupling scales linearly with the NLIW amplitude. At second order (double scattering), where modes n and m interact through an intermediate set of modes j, a three wave resonance condition exists such that wavenumbers other than k_{mn} interact with the modes. Since two scattering events are involved at this order the coupling strength scales quadratically with the NLIW amplitude. The extension of this view to third and higher order scattering is clear. Using direct numerical simulations that include all orders of scattering, previous studies have implicated the beat wavenumber as a significant inverse length scale (Zhou, Zhang, and Rogers, 1991; Preisig and Duda, 1997; Frank *et al.*, 2004; Chiu *et al.*, 2004; Apel *et al.*, 2007), but the present analysis shows mathematically that the beat wavenumber dependence comes primarily out of the single scattering approximation. Interestingly in the second order stochastic analysis of Dozier and Tappert (1978a) the second order resonances collapse to first order ones due to the random phase approximation, and thus only the beat wavenumber is important in that analysis.

Another important and new contribution is that the series expansion reveals significant dependence of energy transfer between modes on the initial conditions. In particular it is shown that the transfer of energy to modes with low or zero initial energy requires a second order scattering process. This result also seems to have not been appreciated in the literature and in recent reviews (Lynch *et al.*, 2005; Apel *et al.*, 2007).

Finally, to delve more deeply into the sensitivity of mode coupling to NLIW space-time scales, a Gaussian ISW model is presented and used with first order theory where the beat wavenumber k_{mn} or the cycle distance $\lambda_{mn}=2\pi/k_{mn}$ plays a critical role. Direct numerical simulation under moderate to strong coupling conditions shows that this model is not grossly inaccurate and that it can be used to reveal the significant parametric dependence of mode coupling on ISW depth structure, horizontal width, and position, and wave packet assemblages, as a function of acoustic frequency. A few of the primary theoretical results are mentioned here.

First, it is seen that there is a sinusoidal modulation of the coupling between modes n and m that depends on the location of the ISW relative to the source: The spatial (temporal) period of this modulation is λ_{mn} ($\lambda_{mn}/c_{\mathrm{ISW}}$), where c_{ISW} is the propagation speed of the ISW. This view of the modulation process provides important physical insight for the interpretation of the quasiregular time modulation of pressure and mode energy patterns seen in observations and models (Duda and Preisig, 1999; Rouseff *et al.*, 2002). In addition it is seen that for assemblages of ISWs there is a double resonance condition that can lead to very strong mode-mode interaction, that is if the individual ISW width and the separation between ISWs in the packet is an integer multiple of cycle distances then the coupling between these modes can be very large. Alternatively if the ISWs are separated by odd multiples of half the cycle distance, the coupling effects will cancel out. This distinction has been discussed by Apel *et al.* (2007) based on numerical simulations, but the present work demonstrates the theoretical foundation of the double resonance. The dependence on the separation of the ISWs is often called the "Bragg" resonance (Zhou, Zhang, and Rogers, 1991; Apel *et al.*, 2007), and is demonstrated in this paper to be in fact distinctly different from the resonance that occurs for individual ISWs in the single scattering approximation.

The outline of this paper is as follows. Section II describes the shallow water acoustic mode equations and their solution and physical interpretation via the Dyson series. In Sec. III single scattering results are used with a Gaussian ISW model, thus yielding analytic formulae for complex modal amplitude; these formulae are used to describe the dependence of mode coupling on ISW parameters. In Sec. IV direct numerical simulations of mode propagation through ISWs demonstrate the accuracy of the first order theory for assemblages and single ISWs, and they reveal the important impact of attenuation induced broadening of mode coupling resonances at low frequency. Final remarks are presented in Sec. V.

II. SHALLOW WATER ACOUSTICS: THE DYSON SERIES

Here the two-dimensional shallow water coupled mode formalism of Creamer (1996) is followed, so only a brief description of the background material is included here. The sound speed field is written $c(r,z)=\bar{c}(z)+\delta c(r,z)$, and the acoustic pressure is written in terms of the unperturbed (i.e., $\delta c=0$) vertical mode functions, $\phi_n(z)$. The pressure has the form

$$p(r,z)=\sum_{n=1}^{N}\frac{A_n(r)}{\sqrt{k_n r}}\phi_n(z), \tag{1}$$

and in this approach the acoustic variability is contained exclusively in the mode amplitude functions $A_n(r)$. In the quasistatic, narrow angle, weak forward scattering approximation Dozier and Tappert (1978a), and Dozier (1983) show

that the scaled mode amplitude $\psi_n = A_n \exp(-il_n r)$ obeys the one-way equation[2]

$$\frac{\partial \psi_n}{\partial r} = -i \sum_{m=1}^{N} \rho_{mn}(r) e^{il_{mn} r} \psi_m(r), \quad l_{mn} = l_m - l_n, \tag{2}$$

where the symmetric coupling matrix ρ_{mn} is given by

$$\rho_{mn}(r) = \frac{k_0^2}{(k_m k_n)^{1/2}} \int_0^\infty \mu(r,z) \frac{\phi_n \phi_m}{\sigma(z)} dz, \quad \mu(r,z) = \frac{\delta c(r,z)}{c_0}. \tag{3}$$

Here the complex horizontal wavenumber is $l_n = k_n + i\alpha_n$, the NLIW sound speed perturbation is $\delta c(r,z)$, the background density profile is $\sigma(z)$, and $k_0 = \omega / c_0$ is a reference wavenumber, ω being the acoustic frequency and c_0 a reference sound speed in the water column.

A. Solution via the Dyson series

Now it is realized that the fractional sound speed perturbation from the NLIWs are relatively small (i.e., $\mu \ll 1$) thus suggesting a perturbation approach. If the coupling matrix ρ_{mn} is of order some small parameter ϵ, then the mode amplitudes $\psi_n(r)$ can be written as a power series of progressive orders of ϵ. Solution of Eq. (2) by iteration then gives the Dyson series (Dyson, 1949; Sakurai, 1985) which was introduced to the ocean acoustics literature by Dozier and Tappert (1978a). Using the Dyson series the change in mode amplitude upto second order is[3]

$$\psi_n(R) = \psi_n(0) - i \sum_{m=1}^{N} \psi_m(0) S_{mn}^{(1)} - \sum_{m=1}^{N} \sum_{j=1}^{N} \psi_j(0) S_{mnj}^{(2)}, \tag{4}$$

where $\psi_n(R)$ is the mode amplitude at the receiver range R, assumed some significant distance from the NLIW, $\psi_n(0)$ is the mode amplitude at the source, and $S_{mn}^{(1)}$ and $S_{mnj}^{(2)}$ are the first and second order Dyson series scattering matrices. These scattering matrices are given by

$$S_{mn}^{(1)} = \int dk \hat{\rho}_{mn}(k) \int_{-R/2}^{R/2} dr' e^{i(l_{mn}-k)r'}, \tag{5}$$

$$S_{mnj}^{(2)} = \int dk \int dk' \hat{\rho}_{mn}(k) \times \hat{\rho}_{jm}(k') \int_{-R/2}^{R/2} dr' \int_{-R/2}^{r'} dr'' e^{i((l_{mn}-k)r' + (l_{jm}-k')r'')}, \tag{6}$$

where $\hat{\rho}_{mn}(k) = 1/2\pi \int_{-\infty}^{\infty} dr \rho_{mn}(r) e^{ikr}$ is the Fourier transform of the coupling matrix.

An important acoustic observable is the change in modal energy after propagation through NLIWs. Multiplying Eq. (4) by its complex conjugate and keeping terms to second order the result is

$$|\psi_n(R)|^2 = |\psi_n(0)|^2 + 2 \sum_{m=1}^{N} \psi_m(0) \psi_n(0) \mathrm{Im}(S_{mn}^{(1)}) - 2 \sum_{m=1}^{N} \sum_{j=1}^{N} \psi_j(0) \psi_n(0) \mathrm{Re}(S_{mnj}^{(2)}) + \sum_{m=1}^{N} \sum_{j=1}^{N} \psi_m(0) \psi_j(0) S_{mn}^{(1)} (S_{jn}^{(1)})^* \tag{7}$$

In the absence of attenuation it can be shown (Dozier and Tappert, 1978a; Dozier, 1983) that because of the symmetry properties of the scattering matrices, Eq. (7) conserves energy at each order;[4] that is to say one need not sum up the whole series to get energy conservation. As would be expected, when attenuation is included, modal energy dissipation depends on the coupling history caused by the NLIW.

1. Sensitivity to NLIW amplitude and horizontal scales

The terms in the Dyson series have important physical meaning, and provide insight into the sensitivity of mode coupling to NLIW amplitude, horizontal spatial scales [as determined by $\hat{\rho}_{mn}(k)$], and the initial conditions $\psi_n(0)$.

The first order term in Eq. (4) physically represents single scattering in which mode n only interacts with mode m. As such the scattering matrix $S_{mn}^{(1)}$ depends linearly on the coupling matrix $[\hat{\rho}_{mn}(k)]$, and thus linearly on the NLIW amplitude. But the scattering matrix gives additional information about the sensitivity of mode coupling to the horizontal spatial scales of the NLIWs. The range integral in Eq. (5) is a filter function on the wavenumber components of the coupling matrix $\hat{\rho}_{mn}(k)$; a simple calculation yields

$$H_{mn}^{(1)}(R,k) = \int_{-R/2}^{R/2} dr' e^{i(l_{mn}-k)r'} = R \, \mathrm{sinc}[(l_{mn}-k)R/2] \tag{8}$$

$$\simeq 2\pi \delta(k_{mn} - k), \quad \alpha = 0, \quad R \to \infty. \tag{9}$$

The filter function, $H_{mn}^{(1)}(R,k)$ shows a resonance at $k = k_{mn} = k_m - k_n$, with a width of $\alpha_{mn} = \alpha_m - \alpha_n$. In the limit of no attenuation and a large range the resonance collapses to a delta function and only the beat wavenumber, k_{mn}, contributes to the coupling. The importance of the beat wavenumber is significant because this wavenumber characterizes the energy carrying envelope of the interfering modes. The fact that attenuation broadens the resonance (a property typical of forced linear systems with damping) is extremely important because it indicates that more NLIW wavenumbers than simply the beat wavenumber contribute to the coupling. While Creamer (1996) has showed that attenuation can dramatically affect the approach to saturation when acoustic propagation through a randomized internal wave field is considered, it is believed that the present result showing directly the effect of attenuation on the scattering matrix and the scales of NLIWs has not been appreciated before and is new. In Sec. IV of this paper numerical examples will be presented to examine the importance of attenuation to the scattering matrix as a function of frequency and position of NLIWs.

The second order terms in Eq. (4) are interpreted as double scattering events in which mode n and m interact through the set j of intermediate modes.[5] Because two scattering events are involved this term depends on the square of the coupling matrix, and thus the scaling of acoustic mode amplitude fluctuations to NLIW amplitude will be quadratic. Again, the sensitivity of mode coupling to NLIW scales at second order is determined by the range integrals in Eq. (6). Here the result is

$$H^{(2)}_{mnj}(k,k',R) = \int_{-R/2}^{R/2} dr' \int_{-R/2}^{r'} dr'' e^{i((l_{mn}-k)r'+(l_{jm}-k')r'')}$$

$$= \frac{R}{i(l_{jm}-k')}(\mathrm{sinc}[(l_{jn}-k-k')R/2]$$

$$- e^{-i(l_{jm}-k')R/2}\mathrm{sinc}[(l_{mn}-k)R/2]), \qquad (10)$$

and it is seen that in addition to the two wave resonance at $k=k_{mn}$, there is a three wave resonance involving the difference wavenumber k_{jn} associated with the intermediate scattering event. Hence, the important conclusion is that in the multiple scattering approximation other NLIW scales besides the beat wavenumber scale will contribute to the coupling. As in the single scattering case these other resonances will be broadened by attenuation.

Previous studies of sound propagation through random fields of internal waves have exclusively attributed mode coupling behavior to the beat wavenumber because of the random phase approximation (Dozier and Tappert, 1978a; Dozier, 1983; Creamer, 1996) and in the case of NLIWs the beat wavenumber has been rather loosely implicated in mode coupling effects (Zhou, Zhang, and Rogers, 1991). Here, it is clearly seen that the beat wavenumber sensitivity is a first order scattering effect (modified by attenuation), and that to higher order there are other wavenumber sensitivities that need to be considered.

2. Mode energy observable: Coupling to weakly excited modes

In addition to demonstrating energy conservation and dissipation principles, The Dyson series mode energy equation [Eq. (7)] reveals the important dependence of mode coupling on the initial mode amplitude distribution $\psi_n(0)$, which is dictated by the source type and depth. In particular, consider the case in which a specific initial mode amplitude is zero [i.e., $\psi_n(0)=0$], then the only nonzero term in Eq. (7) is the fourth term; a second order term. Hence, the important conclusion is that the transfer of energy to a weakly excited mode requires at least a second order scattering process. Second order scattering will only be significant if the scattering matrices are large. Since more typical shallow water point sources distribute energy somewhat more broadly over mode number, the practical first order picture is that mode energy is primarily shared among the modes that initially have the largest energies. Interestingly, the numerical work of Preisig and Duda (1997) utilized a single excited mode, and thus their results involved no first order scattering. Other numerical studies, primarily involved with the interpretation of data, use the more typical point source (Tielburger, Finnette, and Wolf, 1997; Headrick *et al.*, 2000b; Rouseff *et al.*, 2002; Chiu *et al.*, 2004; Frank *et al.*, 2004, 2005; Fredricks *et al.*, 2005). Among these numerical studies Tielburger *et al.* (1997) pointed out the important dependence of mode coupling on initial conditions, so that the present work here reveals the theoretical underpinnings of this dependence.

III. SINGLE SCATTER APPROXIMATION: AN ANALYTIC MODEL

The second order terms in the Dyson series provide important insights into the sensitivity of mode coupling to the scales of NLIWs and to initial conditions, but the primary objective of this work is to delve as deeply as possible into the specific NLIW parameters that affect shallow water acoustic propagation, for example NLIW depth and range structure, range from the source, and wave packet structure. This effort will require a specific model of NLIWs and it will require an acoustic model that is not grossly inaccurate. Therefore, for the remainder of this paper the focus will be on the single scattering approximation, which will allow for a simple first order analytic result.

A. Sound speed perturbations: Gaussian solitons

Consistent with the first order approach a model of internal-wave induced sound-speed perturbations is considered in which the fractional sound-speed change has the form (Munk, 1974)

$$\mu(r,z;t) = -\frac{1}{c_0}\left(\frac{dc}{dz}\right)_p \zeta(r,z;t) = -\frac{1}{c_0}\left(\frac{dc}{dz}\right)_p \zeta_0 W_1(z)F(r,t), \qquad (11)$$

where $(dc/dz)_p$ is the potential sound-speed gradient, and $\zeta(r,z;t)=\zeta_0 W_1(z)F(r;t)$ is the internal-wave vertical displacement. In this treatment the time variable t is considered a parameter since it is assumed that the sound speed field is frozen in time during the propagation of the acoustic signal (frozen field approximation). Consistent with observations of NLIWs (Liu, 1988; Apel *et al.* 2007, 1997; Colosi *et al.* 2001; Duda *et al.*, 2004a; Liu *et al.*, 2004), a mode 1 structure is assumed given by $W_1(z)$ and the mode is normalized such that its maximum amplitude is unity. The parameter ζ_0, therefore gives the maximum displacement of the internal wave (the amplitude), and the function $F(r;t)$ describes the range and time variation of the NLIW. This approximation, in which μ factors into separate depth and range functions has been used by other investigators (Chiu *et al.*, 2004; Badiey *et al.*, 2007) but can break down if the internal wave displacements or the thermocline gradients are too large (Preisig and Duda, 1997). In practice, it must be pointed out, that this approximation need not be made as the coupling matrix $\rho_{mn}(r)$ can be evaluated for any functional form of $\mu(r,z)$; the factorized form is used as a first approximation to gain physical insight. Thus, the mode coupling matrix and its Fourier transform become

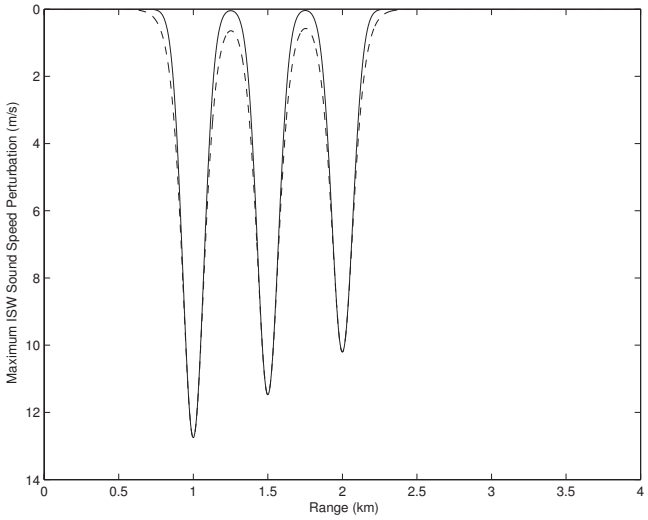


FIG. 1. Range variation of maximum sound speed perturbations from a packet of ISWs modeled as Gaussians (solid) and hyperbolic secant squared functions (dash). The ISWs all have a range width of $\Delta = 100$ m.

$$\rho_{mn}(r;t) = \frac{k_0^2}{(k_m k_n)^{1/2}} \int_0^\infty \mu(r,z;t) \frac{\phi_n(z)\phi_m(z)}{\sigma(z)} dz = Z_{mn} F(r;t), \tag{12}$$

$$\hat{\rho}_{mn}(k;t) = \frac{Z_{mn}}{2\pi} \int_{-\infty}^\infty dr F(r;t) e^{ikr}, \tag{13}$$

where

$$Z_{mn} = -\frac{\zeta_0}{c_0} \frac{k_0^2}{(k_m k_n)^{1/2}} \int_0^\infty \left(\frac{dc}{dz}\right)_p (z) W_1(z) \frac{\phi_n(z)\phi_m(z)}{\sigma(z)} dz. \tag{14}$$

Here, the matrix Z_{mn} represents the contribution to mode coupling from the depth structure of the acoustic modes and the NLIW.

A simple analytic model that demonstrates the scattering physics very well comes about by considering a "Gaussian" soliton to represent the range variation of the NLIW. That is to say the approximation is made that each NLIW is an ISW and is of the form

$$F(r;t) = \mathrm{sech}^2\left[\frac{(r - r_0(t))}{\Delta}\right] \simeq \exp\left[-\frac{(r - r_0(t))^2}{\Delta^2}\right], \tag{15}$$

where $r_0(t)$ and Δ are the position (a function of time because of ISW propagation) and the width of the ISW. Figure 1 shows an example of a three ISW packet, and the Gaussian approximation is demonstrated to be excellent. Also, numerical work has shown that the acoustic mode coupling properties of the Gaussian ISW are extremely similar to those of the sech-squared ISW (Young, 2006). In this simple model, however, an often important feature of NLIWs is ignored, namely the bore-like structure that can occur near internal tide generation regions (Colosi *et al.*, 2001). This type of feature can be treated in the general theory but there is no simple analytic model.

B. Mode energy fluctuations

With the Gaussian sound speed perturbation model, an expression for mode energy fluctuations can be written down. The scattering matrix for the Gaussian soliton is written

$$S_{mn}^{(1)} = Z_{mn} \frac{\Delta}{2\pi^{1/2}} \int_{-\infty}^\infty dk e^{ikr_0(t)} \exp\left[-\frac{k^2\Delta^2}{4}\right] H_{mn}(R,k). \tag{16}$$

An analytic solution for $S_{mn}^{(1)}$ is not available for nonzero α_n, but at higher acoustic frequencies where bottom attenuation is weak, the effects of attenuation on the scattering matrix $S_{mn}^{(1)}$ are expected to be small (as will be demonstrated in Sec. IV). Hence, in this approximation the filter function $H_{mn}^{(1)}$ takes the particularly simple form of a delta function [i.e., Eq. (9)]. Using Eq. (7), and converting to mode amplitude A_n, the mode energy is given by

$$|A_n(R;t)|^2 = \left[|A_n(0)|^2 + 2\pi^{1/2} \sum_{m=1}^N A_{mn}(0) Z_{mn} \Delta \times \exp\left[-\frac{k_{mn}^2\Delta^2}{4}\right] \sin(k_{mn} r_0(t))\right] e^{-2\alpha_n R}, \tag{17}$$

where $A_{mn}(0) = A_m(0)A_n(0)$ is a matrix of the initial mode amplitudes. This result is a good starting place to discuss the sensitivity of mode coupling to the various parameters of the problem.

1. Dependence on ISW displacement, ζ_0

As previously discussed in Sec. II A in the single scattering approximation the fluctuation in mode amplitude will be a linear function of the ISW amplitude ζ_0, and this scaling also holds for the mode energy $|A_n|^2$ as is demonstrated by Eqs. (17) and (14). One further comment is warranted on this subject. The linear scaling with ζ_0 provides an important distinction between the mode coupling that occurs from ISW waves of depression ($\zeta_0 < 0$) and waves of elevation ($\zeta_0 > 0$). Namely, the only difference between the acoustical effects of waves of elevation and depression is the direction in which energy is transferred between mode n and m.

2. Dependence on ISW width, Δ

The dependence of mode coupling on the horizontal structure of ISWs in the single scattering approximation is completely contained in the nature of the coupling matrix at the beat wavenumber $\hat{\rho}_{mn}(k = k_{mn})$, however, it is useful to view this effect in the spatial domain through the ISW width parameter Δ. Equation (17) shows that the mode energy fluctuation is controlled by the term $\Delta \exp(-k_{mn}^2\Delta^2/4)$, and thus in the limit of both large and small Δ the coupling vanishes, as expected. Maximum coupling between modes m and n occurs when the full width of the ISW (i.e., 2Δ) approximately matches the cycle distance of the beat pattern of these modes (i.e. $\lambda_{mn} = 2\pi/k_{mn}$); for the Gaussian ISW the result is

$$(2\Delta)(\text{Max coupling}) = \frac{2^{1/2}}{\pi}\lambda_{mn}. \quad (18)$$

It is important to note also that in any geographic location in the ocean there will be a distribution of ISWs with different widths (Colosi *et al.*, 2001; Duda *et al.*, 2004a; Ramp *et al.*, 2004), so that the response of mode coupling to changing Δ is an important issue. Near the maximum response, a doubling/halving of the ISW width will result in a 44%/72% reduction in the mode energy change.

Because of the dependence on the beat wavenumber, the acoustic sensitivity to ISW width depends critically on the dispersion relation, which in turn depends on the acoustic frequency, water depth, and density and sound speed profiles. Some general comments can be made about the dependence on acoustic frequency and water depth. In particular as frequency and water depth increase the separation between the eigenwavenumbers decreases; this means λ_{mn} increases. Hence, if either the frequency or the water depth increase the acoustic field will be more sensitive to ISW's with larger Δ. This consideration is important especially in acoustic transmissions near the continental slope (Fredricks *et al.* 2005; Chiu *et al.* 2004). Another consideration is the shape of the dispersion curve. If the form is nearly linear then many mode combinations will have similar k_{mn} values and thus will have similar contributions to the coupling from the ISW width term. If, however, the dispersion relation is non-linear (as is more typical), there will be a greater diversity of k_{mn} values and thus potentially a strong mode number dependence to the coupling from this ISW width term.

3. Dependence on ISW position, $r_0(t)$

One of the most notable features of early observations and modeling studies of sound transmission through NLIWs was the quasiregular time dependence of the acoustic field as the NLIWs propagate through the acoustic path (Duda and Preisig, 1999; Rouseff *et al.*, 2002). The present model shows precisely the cause of this time modulation, which is the relative phasing of modes n and m at the location of the ISW, $r_0(t)$. The mode coupling channels in Eq. (17) are seen to be sinusoidally modulated through the term $\sin(k_{mn}r_0(t))$, thus the temporal period of this modulation will be the time it takes the ISW to propagate a cycle distance or $\lambda_{mn}/c_{\text{ISW}}$. Since in general mode n will interact with several other neighboring modes, the time dependence of mode n energy will depend on the different modulation times of the modes it interacts with. This single scattering result is the theoretical basis for the quasiregular time dependence of acoustic fields examined by previous studies (see, for example, Rouseff *et al.*, 2002, Figs. 8 and 14, as well as Duda and Preisig, 1999, Fig. 2).

Again, because the modulation of mode coupling due to the ISW location r_0 is driven by the beat wavenumber, there are important dependencies, on the shape of the dispersion relation, acoustic frequency, and water depth. As discussed previously increasing acoustic frequency and/or increasing water depth will increase the cycle distance λ_{mn}, hence, the impact of these changes is an increase in the modulation time, an important factor for coherent processing over time. In addition, a nearly linear dispersion relation would give rise to nearly commensurate modulation frequencies and thus long timescale periodicities in the overall mode energy variability; a nonlinear dispersion rule on the other hand would have noncommensurate frequencies and thus no long timescale periodicities.

4. Dependence on ISW depth structure, Z_{mn}

To examine the sensitivity to ISW depth structure, a simplified model is considered in which there is a constant background sound speed c_0, buoyancy frequency N_0, and potential sound speed gradient. In this case both the acoustic and internal wave modes are sinusoids and the depth matrix takes the form (Creamer, 1996)

$$Z_{mn} = -\frac{\zeta_0}{c_0}\frac{k_0^2}{(k_m k_n)^{1/2}}\left(\frac{dc}{dz}\right)_p \Upsilon_{mn}, \quad (19)$$

$$\Upsilon_{mn} = \frac{2j_{\text{ISW}}}{(m-n)^2 - j_{\text{ISW}}^2} \quad m+n+j_{\text{ISW}} \text{ odd},$$

$$= -\frac{2j_{\text{ISW}}}{(m+n+1)^2 - j_{\text{ISW}}^2} \quad m+n+j_{\text{ISW}} \text{ even}. \quad (20)$$

In Eq. (20) the ISW mode number j_{ISW} acts like a roll-off parameter, and thus only for mode number separations greater than j_{ISW} do the matrix elements of Υ_{mn} start to decay. Observationally it has been well established that shallow water ISWs have a mode 1 vertical structure (Apel *et al.*, 2007) so that $j_{\text{ISW}}=1$. Hence, it is seen that for typical ocean conditions the off diagonal matrix elements decay rapidly (i.e., roughly as the mode number separation to the -2 power), and this decay strongly enforces the dominance of near neighbor coupling. It should also be noted that no other terms in the mode energy equation [Eq. (17)] have this strong dependence on mode number separation, and thus near neighbor coupling is determined by ISW depth structure. The relevance of ISW depth structure on near neighbor mode coupling has been pointed out previously by Chiu *et al.* (2004) with regards to direct numerical simulations of South China Sea propagation, thus giving credence to the simple sinusoidal calculation done here.

5. Dependence on ISW packets

It is often the case that multiple ISWs are present along an acoustic propagation path (Apel *et al.*, 2007), and thus the situation of multiple scattering from assemblages of ISWs is significant (Zhou, Zhang, and Rogers, 1991). Because of nonlinear dynamics, multiply interacting ISWs do not follow the superposition principle, but to have a nonlinear interaction two (or more) ISWs must be in close proximity. Observationally the case of nonlinear interacting ISWs seems to be more the exception than the rule (Colosi *et al.*, 2001; Duda *et al.*, 2004a; Apel *et al.*, 2007) and thus in the case in which the ISWs are sufficiently isolated superposition may be used yielding

$$\mu(r,z) = \sum_{p=1}^{N_s} \mu_p(r,z), \tag{21}$$

where μ_p is the fractional sound speed perturbation of the pth ISW, and N_s is the number of ISWs. Therefore, in the single scattering approximation the total scattering matrix is just the sum of all the individual scattering matrices and becomes

$$S_{mn}^{(1)}(R) = \sum_{p=1}^{N_s} \int dk \hat{\rho}_{mn}(k,p) H_{mn}(R,k)$$
$$= \pi^{1/2} \sum_{p=1}^{N_s} Z_{mn}(p) \Delta_p \exp\left[-\frac{k_{mn}^2 \Delta_p^2}{4} \right] e^{ik_{mn} r_0(t,p)}, \tag{22}$$

where Δ_p, and $r_0(t,p)$ are the width and position of the pth ISW, and Z_{mn} has a p dependence due to $\zeta_0(p)$, the amplitude of the ISW. An important picture emerges from this treatment of the mode coupling, that is mode coupling from N_s ISWs in the single scattering approximation does not necessarily translate into a linear increase in mode energy fluctuation with increasing N_s. The positioning of the ISWs, $r_0(t,p)$, is critical in this regard because of the phase term in the scattering matrix. All other things being equal, if the ISWs are separated by an integer multiple of cycle distances then the energy fluctuation between modes n and m due to the two ISWs adds constructively, if they are separated by an odd multiple of a half cycle they cancel. One then may rightly consider an assemblage of ISWs, each separated by an integer number of cycle distances, as a strongly resonant assemblage for the modes n and m; this resonance is termed a "Bragg" resonance and was discussed by Zhou, Zhang, and Rogers (1991), and Apel *et al.* (2007). Further, if the widths of the ISWs in the assemblage are also nearly a cycle distance wide (see Sec. III B 2) then this situation would be doubly resonant.

Because of the cycle distance dependence on coupling from an assemblage of ISWs, the frequency dependence (and water depth dependence) can be understood. Again as in Secs. III B 2 and III B 3 as frequency and/or water depth are increased cycle distance is increased, meaning that the most resonant assemblages of ISWs would be further apart.

6. Other dependence on acoustic frequency

The acoustic frequency dependence on terms in the mode energy equation related to ISW width, position, and ISW number have been discussed but acoustic frequency enters in a few other terms of note, namely Z_{mn} and $A_{mn}(0)$. In the depth matrix Z_{mn} the primary frequency dependence is from the wavenumber ratio $k_0^2/(k_n k_m)^{1/2}$ which scales linearly with frequency, thus the coupling strength contribution from here would increase linearly with frequency. Because of the effects of the other terms one would not expect the coupling strength to increase without bound. The matrix of initial mode amplitudes $A_{mn}(0)$ has a frequency dependence that is non-trivial since various acoustic modes at different frequencies can have nulls or maxima at the source depth.

IV. A NUMERICAL EXPERIMENT

Direct numerical simulations of multiple scattering mode propagation through ISWs are now presented to compare to the single scattering analytic predictions. These comparisons serve a few purposes. First, they will demonstrate the accuracy of the single scattering approximation for both single and multiple ISW packets. The numerical simulations will also be used to evaluate the important effects of attenuation on the scattering matrix $S_{mn}^{(1)}$. From Sec. II A 1 it was seen that attenuation broadens the wavenumber resonance at $k=k_{mn}$, and numerical calculations are required to determine if the broadening is observationally significant at various acoustic frequencies.

A. The simulation

For the numerical calculations, an analytic background sound speed profile inspired by Duda and Preisig (1999) is chosen of the form

$$\bar{c}(z) = c_s - \frac{\Delta c}{2}\left[1 + \tanh\left(\frac{z - z_{\text{th}}}{\Delta_{\text{th}}} \right) \right], \quad 0 \leqslant z \leqslant D = 100 \text{ m},$$
$$= c_{\text{bot}}, \quad 100 \text{ m} < z, \tag{23}$$

where D is the water depth, $c_s = 1520$ m/s is the surface sound speed, $\Delta c = 40$ m/s is the change in sound speed through the thermocline, $z_{\text{th}} = 40$ m is the thermocline depth, and $\Delta_{\text{th}} = 15$ m is the thermocline width. In order to better demonstrate the effects of seafloor loss on mode coupling a seafloor with large attenuation is selected. Parameters are chosen for a sandy seafloor giving $c_{\text{bot}} = 1650$ m/s, and $\alpha^{\lambda} = 0.8$ dB/λ (Jensen *et al.*, 1993). For these parameters a limiting bottom grazing angle of 26.2° is realized. For density variations, the watercolumn and seafloor are modeled with constant densities of $\sigma = 1000$ kg/m^3 and $\sigma = 1900$ kg/m^3, respectively.

The acoustic propagation simulations are carried out via numerical solution of Eq. (2) as described in Dozier and Tappert (1978b), and Creamer (1996). The final propagation range is $R = 22$ km, and a point source is simulated at frequencies of 100 and 400 Hz, for a source depth is 90 m. The mode functions $\phi_n(z)$ and the eigenwavenumbers k_n are generated in the standard way (Jensen *et al.*, 1993), and the imaginary part of the modal wavenumber α_n (modal attenuation) is computed using perturbation theory (Jensen, *et al.*, 1993). A few of the modal attenuation values are given by

$$\alpha_1 = 0.0361, \quad \alpha_2 = 0.0791, \quad \alpha_3 = 0.0896,$$
$$\alpha_4 = 0.1369, \quad \alpha_5 = 0.2256, \quad \alpha_6 = 0.7090, \ldots \text{ (rad/km)},$$

for 100-Hz, and

$$\alpha_1 = 0.0099, \quad \alpha_2 = 0.0259, \quad \alpha_3 = 0.0430,$$
$$\alpha_4 = 0.0591, \quad \alpha_5 = 0.0727, \quad \alpha_6 = 0.0818, \ldots \text{ (rad/km)},$$

for 400 Hz. The mode calculation naturally leads to trapped and untrapped modes for consideration, and somewhat arbitrarily an equal number of each type of mode is included in

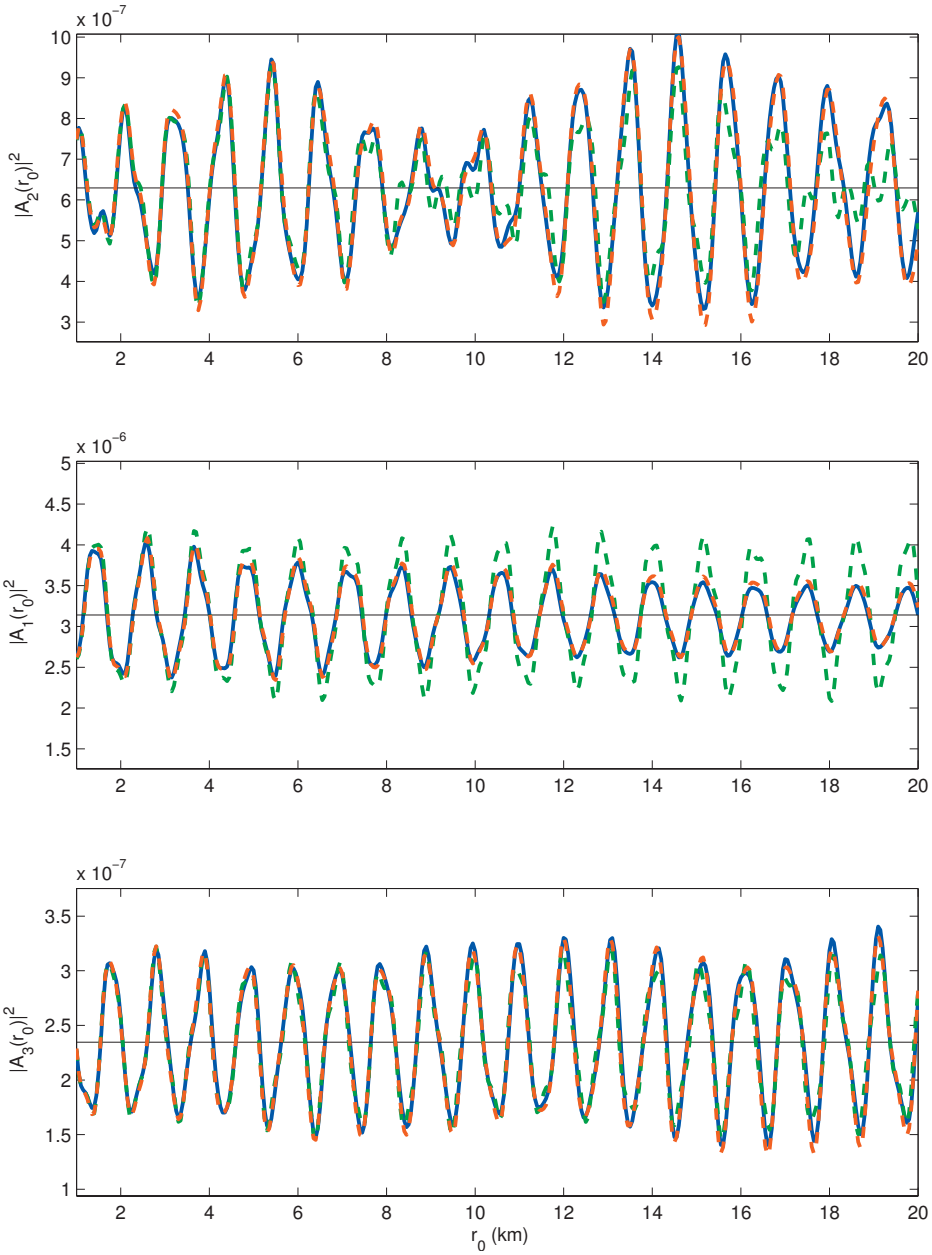


FIG. 2. Direct numerical simulation (blue solid), and theoretical results for 100 Hz mode energy for modes 2, 1, and 3, as a function of a single soliton position r_0. The bottom attenuation is $\alpha^{\lambda}=0.8$ dB/λ. Theoretical results in which the scattering matrix incorporates the effects of attenuation [Eq. (28)] are plotted with red dash lines, while the no attenuation results [Eq. (27)] are plotted with a green dash. The solid horizontal lines give the value of the unperturbed mode energy.

the simulation. At 100 Hz, six trapped and untrapped modes are used for a total of $N=12$, while at 400 Hz, 22 trapped and untrapped modes yield a total of $N=44$.

The fractional sound speed fluctuation μ is modeled as a train of Gaussian ISWs using

$$\mu(r,z) = -\frac{1}{c_0}\frac{d\bar{c}}{dz}\sin(\pi z/D)\sum_{p=1}^{N_s}\zeta_0(j)F(r,\Delta_p,r_0(p)),\ z \leqslant D, \tag{24}$$

with $\mu=0$ in the seabed. Here, the potential gradient of sound speed is approximated using the total gradient (a good assumption in the upper ocean), and a sinusoidal shape to the depth structure of the ISWs is assumed. The factors $\zeta_0(p)$ are the vertical displacements of the various ISWs, which are chosen to be waves of depression with $\zeta_0=10$ m for a single wave case and $\zeta_0=10$, 9, and 8 m for a three ISW packet. The width of the ISWs Δ_p are all chosen to be 100 m, and their relative positions are 500 m apart with the largest one occurring nearest the source. These parameter choices are consistent with observations made on the New England Shelf (Colosi *et al.*, 2001) and lead to a maximum sound speed change for each of the ISWs of 12.75, 11.48, and 10.2 m/s at a depth of 41 m; a change roughly 1/3 the total waveguide deviation of 40 m/s. Figure 1 shows an example of the range variation of maximum sound speed perturbation from the three ISW packet. Finally a range of ISW positions, r_0, from 1 to 20 km are examined in the calculations.

B. Acoustic observables

In this numerical experiment, the acoustic observables of modal energy, and phase fluctuation are examined, since they quantify the basic nature of the complex mode amplitude at the receiver. For both these observables single scattering theoretical predictions are presented. For modal en-

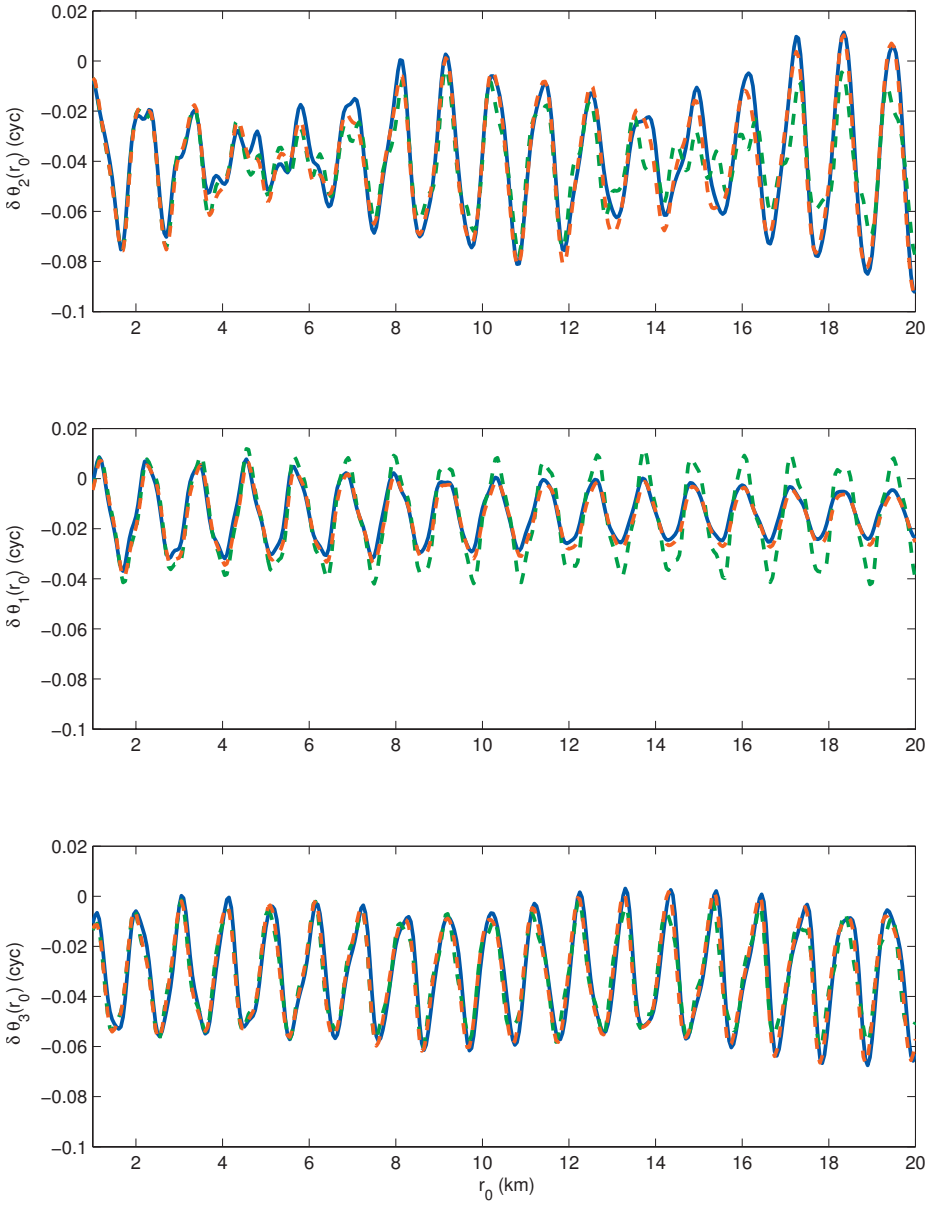


FIG. 3. Same as Fig. 2, except modal phase fluctuation is presented.

ergy which has already been discussed in detail, our theoretical calculations will utilize Eq. (7) truncated to first order, and converting from ψ_n to A_n we obtain

$$|A_n(R)|^2 = \left(|A_n(0)|^2 + 2\sum_{m=1}^{N} A_{mn}(0)\mathrm{Im}(S_{mn}^{(1)}) \right) e^{-2\alpha_n R}. \quad (25)$$

For phase fluctuations $\delta\theta_n$, Eq. (4) is truncated to first order to give

$$\delta\theta_n = \mathrm{Arg}\left(\psi_n(0) - i\sum_{m=1}^{N} \psi_m(0) S_{mn}^{(1)} \right). \quad (26)$$

Note that in the equation for ψ_n the deterministic phase is removed and thus only the ISW induced phase effect remains. For these observables we consider two forms of the scattering matrix $S_{mn}^{(1)}$ for assemblages of Gaussian ISWs; these two forms either include or exclude the effects of attenuation and are based on the filter functions $H_{mn}^{(1)}$ given by Eqs. (8) and (9). In particular we use

$$S_{mn}^{(1)} = \pi^{1/2} \sum_{p=1}^{N_s} Z_{mn}(p)\Delta_p \times \exp\left[-\frac{k_{mn}^2 \Delta_p^2}{4} \right] e^{ik_{mn} r_0(p)} \quad \text{no attenuation}, \quad (27)$$

$$S_{mn}^{(1)} = \sum_{p=1}^{N_s} \frac{Z_{mn}(p)}{2\pi^{1/2}} \Delta_p \int dk \exp\left[-\frac{k^2 \Delta_p^2}{4} \right] e^{ikr_0(p)} R \times \mathrm{sinc}[(l_{mn} - k)R/2] \quad \text{attenuation included}. \quad (28)$$

In Eq. (27) we see that the delta function behavior of the resonance says that only one wavenumber contributes to the coupling; numerical tests using the sinc form of the filter function for $\alpha = 0$ and finite range R show no difference from the delta function result.

As previously mentioned, full physics, multiple scattering calculations were carried out by numerical solution of Eq. (2). In these calculations attenuation is always included.

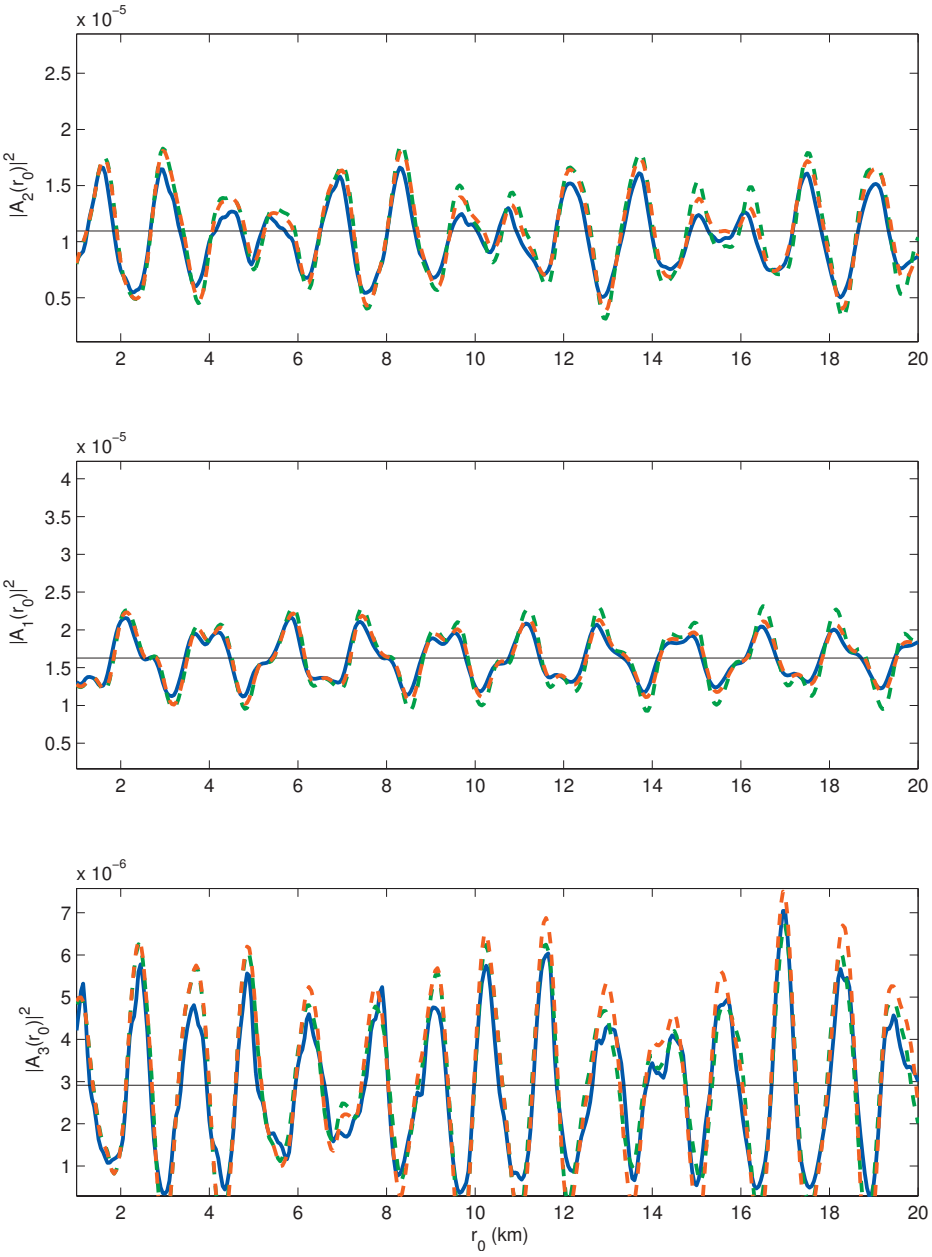


FIG. 4. Direct numerical simulation (blue solid), and theoretical results for 400 Hz mode energy for modes 2, 1, and 3 as a function of a single soliton position r_0. The bottom attenuation is $\alpha^{\lambda}=0.8$ dB/λ. Theoretical results in which the scattering matrix incorporates the effects of attenuation [Eq. (28)] are plotted with red dash lines, while the no attenuation results [Eq. (27)] are plotted with a green dash. The solid horizontal lines give the value of the unperturbed mode energy.

Mode energy and phase fluctuation are then derived from the full physics calculation and these results are compared to the single scattering theoretical results just discussed. The comparisons are presented in the following sections.

C. Single ISW results

Figures 2–5 show comparisons of the direct numerical simulation to the theory, as a function of a single ISW position r_0 for the frequencies of 100 (Figs. 2 and 3) and 400 Hz (Figs. 3 and 4). In these figures mode numbers 2, 1, and 3 are shown because they have the largest initial mode energies. Also in these figures, direct numerical simulation results are plotted with a solid blue line, while analytic results with/without attenuation effects in the scattering matrix are plotted with red/green dash lines. Because the results are plotted as a function of r_0 the quasiregular modulation of the energy fluctuation is readily apparent.

The effects of attenuation on the scattering matrix are discussed first. At the lower frequency, where modal attenuation is largest, it is seen that attenuation dependence in the scattering matrix is critically important for accurately predicting the modal energy and phase fluctuation, particularly for modes 1 and 2. In addition, the effect of attenuation on the mode fluctuations is much larger when the ISW is further from the source; note that all three curves in Figs. 2 and 3 are quite close for smaller values of r_0. At the higher frequency of 400 Hz the difference between the theoretical predictions with and without attenuation are much smaller, and they differ from the direct numerical simulation because of missing multiple scattering physics in the single scattering approximation. Here it is concluded that at 400 Hz, attenuation plays a small role in modifying the mode coupling.

Regarding the accuracy of the single scattering approximation, Figs. 2–5 show that the theory works extremely well at 100 Hz, with deviations of order a few percent, while at

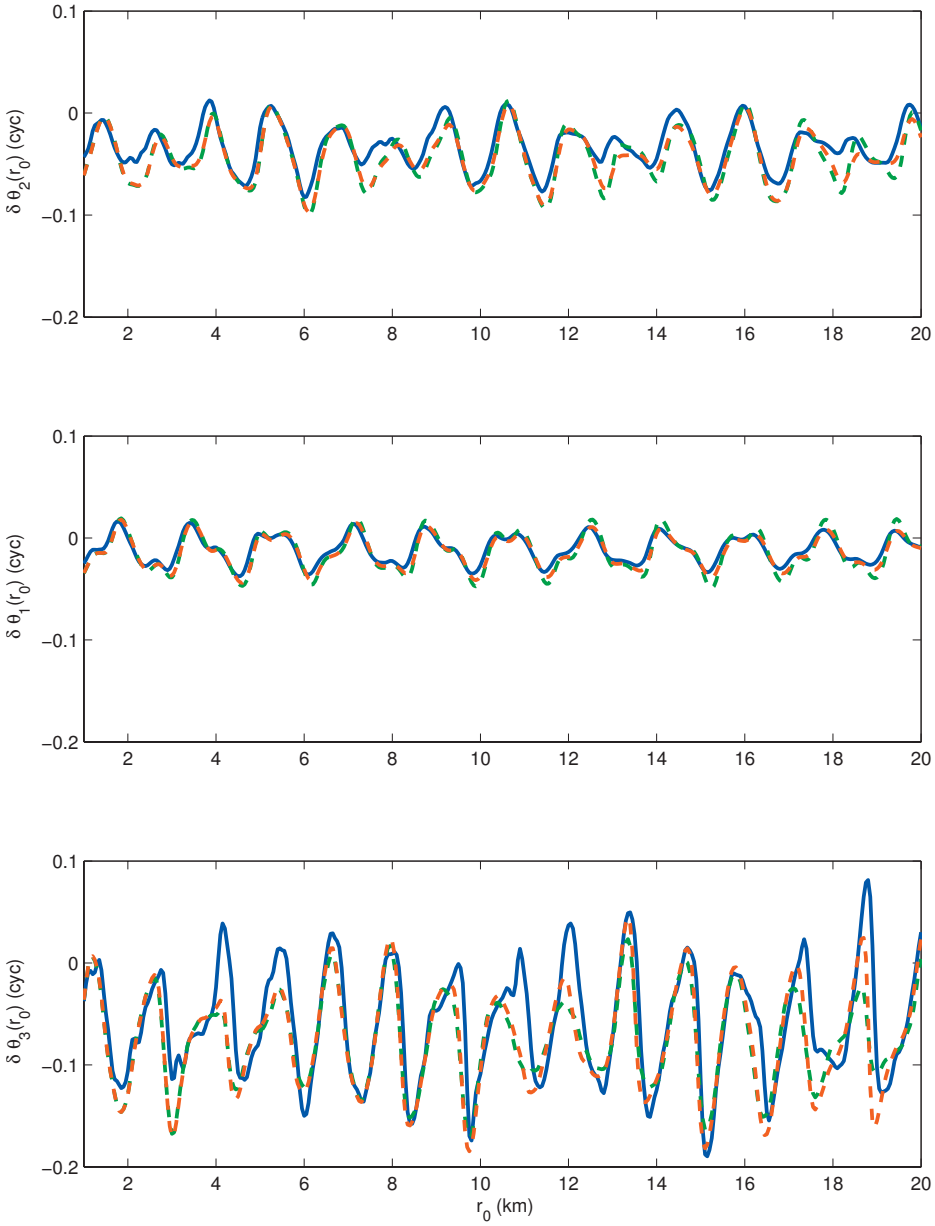


FIG. 5. Same as Fig. 4, except modal phase fluctuation is presented.

400 Hz the theory is working relatively well with maximum deviations of order 10%. In particular the 400-Hz results are striking since the relative change in mode energy is as large as 100% for mode 3, and 50% for modes 1 and 2. Note also that the theory accurately predicts the mean decrease in phase at both frequencies, due to the fact that the ISW produces an overall increase in sound speed (decrease in travel time) in the water column.

D. Multiple ISW results

Next the effects of multiple soliton packets are examined, utilizing the three ISW model; Figs. 6 and 7 show the same conditions as Figs. 2 and 4 except now a three ISW packet is modeled. In these figures the parameter r_0 gives the position of the first, largest ISW, and only analytic results with attenuation in the scattering matrix are plotted. As before, the results at 100 Hz are excellent. At 400 Hz some degradation of the accuracy of the theory is apparent, especially for mode 3. This degradation is expected as the scattering will be more complex in an assemblage of ISWs. However, it must be noted that the mode energy fluctuations for the three ISW packet show some surprising differences from the single ISW cases. At 100 Hz the effect of having three ISWs is to somewhat increase the fluctuations and make the patterns of variability as a function of r_0 somewhat more complex (see Figs. 2 and 6). However, at 400 Hz, the three ISW packet has the surprising effect of reducing the fluctuations relative to the one ISW case (see Figs. 4 and 7). Thus as described in Sec. III B 5 the relationship between mode fluctuations and the number of ISWs depends critically on the relative locations of the ISWs in the packet and the frequency dependent mode cycle distances.

While these calculations only represent a few particular cases, they suggest that single scattering theory can accurately predict complex mode fluctuations (amplitude and phase) in conditions of moderate to strong sound speed perturbations and for acoustic frequencies of order several hundreds of Hz. Previous theoretical work has focused entirely

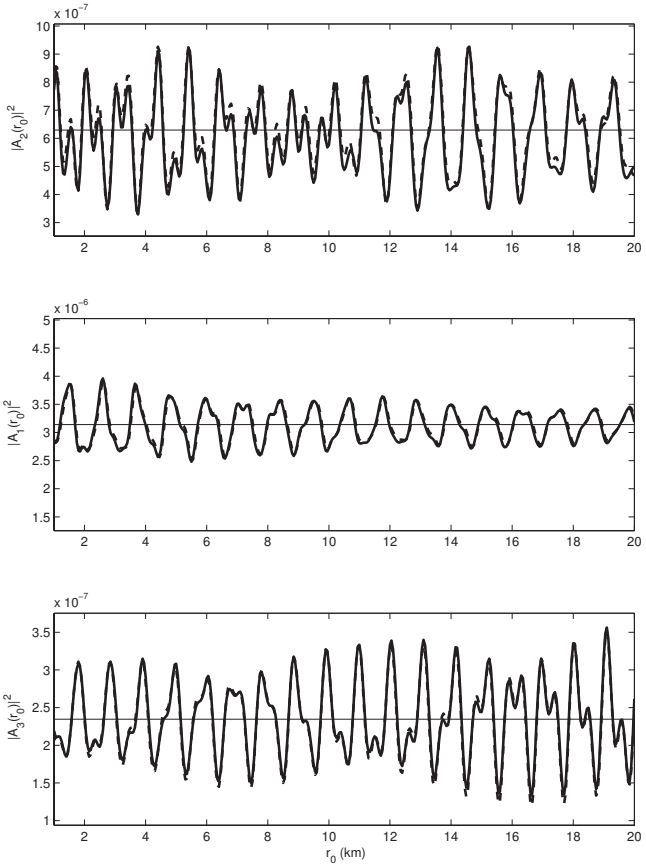


FIG. 6. Direct numerical simulation (solid), and theoretical results (dash) for 100 Hz mode energy as a function of a three soliton wavepacket location r_0. The bottom attenuation is $\alpha^{\lambda}=0.8$ dB/λ. The theoretical results incorporate the effect of attenuation on the scattering matrix [Eq. (28)]. Here, r_0 corresponds to the location of the first and largest soliton in the packet. The solid horizontal lines give the value of the unperturbed mode energy.

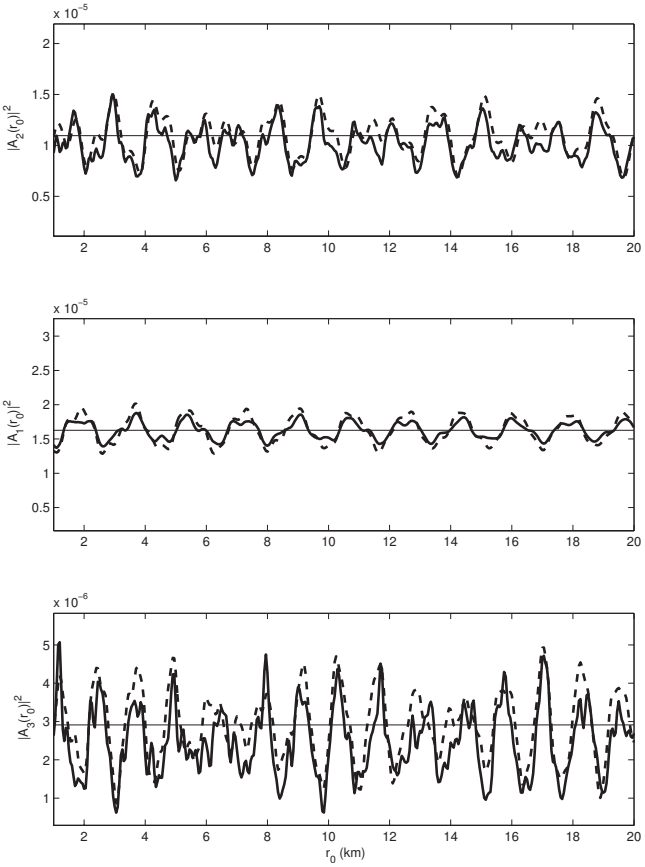


FIG. 7. Same as Fig. 6 except 400-Hz results are presented.

on mode energy (Dozier and Tappert, 1978a; Creamer, 1996; Presig and Duda, 1997), but in the present study good agreement is found on the matter of both phase and energy. The phase effects will be particularly important regarding temporal coherence as discussed by Rouseff *et al.* (2002). These results give some cause for optimism in the efforts to predict acoustic fluctuations caused by NLIWs, a view that contrasts to the rather pessimistic spirit in the literature that considers numerical models as the only tool to study this problem.

V. CONCLUDING REMARKS

In this paper an analytic model of sound propagation through NLIWs is presented which demonstrates much of the important physics of coupled mode propagation in the shallow water environment in which the acoustic path is nearly at right angles to NLIW fronts. The analytic results reveal the sensitivity of mode coupling to the background ocean conditions (particularly attenuation), and the specific space-time parameters of the shallow water NLIW field, as a function of acoustic frequency. Comparisons to direct numerical simulations show that the simplified analytic results are robust in moderately strong coupled mode propagation conditions, showing that the theory indeed provides a first order description of shallow water propagation through NLIWs. The need now is to apply these results to various experimental situations.

ACKNOWLEDGMENTS

The basis for much of this paper was part of a Naval Postgraduate School Oceanography/Meteorology Masters Thesis written by Lieutenant Aaron Young of the Australian Navy. Aaron's excellent work was the first demonstration of the utility of this model. The author is also grateful to Jim Lynch, Boris Ketznelson, Tim Duda and Andrey Morozov for providing very useful comments on the topics of this paper. Allan Pierce and three anonymous reviewers provided excellent guidance on the revision of this manuscript.

[1]Dozier and Tappert (1978a) do not explicitly attribute their solution to Dyson.

[2]It should be noted that Eq. (2) admits wider-angle propagation than the standard parabolic equation (SPE); in the SPE approximation the factor $(k_m k_n)^{1/2}$ in Eq. (3)is replaced by k_0 and the wavenumber is $l_n=(k_n^2-k_0^2)/2k_0+i\alpha_n$. The extent to which Eq. (2) and the symmetry properties of the coupling matrix correspond to present shallow-water, wide-angle propagation codes is an open scientific question.

[3]Also see Dozier and Tappert (1978a) Eq. (45).

[4]Energy conservation means $\Sigma_{n=1}^{N}|\psi_n(R)|^2=\Sigma_{n=1}^{N}|\psi_n(0)|^2$.

[5]In their numerical work, Preisig and Duda (1997), in fact, show the effect of the intermediate states as the modes couple through the NLIWs.

Apel, J., Ostrovsky, L., Stepanyants, Y., and Lynch, J. (**2007**). "Internal solitons in the ocean and their effects on underwater sound," J. Acoust. Soc. Am. **121**, 695–722.

Apel, J., Badiey, M., Chiu, C. S., Finette, S., Headrick, R., Kemp, J., Lynch, J., Newhall, A., Orr, M., Pasewark, B., Teilbuerger, D., Turgut, A., von der Heydt, K., and Wolf, S. (**1997**). "An overview of the 1995 SWARM shallow water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–500.

Badiey, M., Katsnelson, B., Lynch, J., Pereselkov, S., and Siegmann, W. (**2005**). "Measurement and modeling of three-dimensional sound intensity variations due to shallow water internal waves," J. Acoust. Soc. Am. **117**, 613–625.

Badiey, M., Katsnelson, B., Lynch, J., and Pereselkov, S. (**2007**). "Frequency dependence and intensity fluctuations due to shallow water internal waves," J. Acoust. Soc. Am. **122**, 747–760.

Chiu, C. S., Ramp, S., Miller, C., Lynch, J., Duda, T., and Tang, T. (**2004**). "Acoustic intensity fluctuations induced by South China Sea internal tides and solitons," IEEE J. Ocean. Eng. **29**, 1249–1262.

Colosi, J. A., Beardsley, R. C., Gawarkiewicz, G., Lynch, J. F., Chiu, C., and Scotti, A. (**2001**). "Observations of nonlinear internal waves on the outer New England continental shelf during the summer shelfbreak PRIMER study," J. Geophys. Res., [Oceans] **106**, 9587–9601.

Creamer, D. B. (**1996**). "Scintillating shallow water waveguides," J. Acoust. Soc. Am. **99**, 2825–2838.

Dozier, L. B., and Tappert, F. D. (**1978a**). "Statistics of normal-mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **63**, 353–365.

Dozier, L. B., and Tappert, F. D. (**1978b**). "Statistics of normal-mode amplitudes in a random ocean. II. Computations," J. Acoust. Soc. Am. **64**, 353–365.

Dozier, L. B. (**1983**). "A coupled mode model for spatial coherence of bottom-interacting energy," in *Proceedings of the Stochastic Modeling Workshop*, edited by Spofford, C. W., and Haynes, J. M. (ARL-University of Texas, Austin Texas).

Duda, T., and Preisig, J. (**1999**). "A modeling study of acoustic propagation through moving shallow water solitary wave packets," IEEE J. Ocean. Eng. **24**, 16–32.

Duda, T., Lynch, J., Irish, J., Beardsley, R., Ramp, S., Chiu, C. S., Tang, T., and Yang, Y. (**2004a**). "Internal tide and nonlinear internal wave behavior at the continental slope in the Northern South China Sea," IEEE J. Ocean. Eng. **29**, 1105–1130.

Duda, T., Lynch, J., Newhall, A., Wu, L., and Chiu, C. S. (**2004b**). "Fluctuations of 400 Hz sound intensity in the 2001 ASIAEX South China Sea experiment," IEEE J. Ocean. Eng. **29**, 1264–1279.

Dyson, F. J. (**1949**). "The radiation theories of Tomonaga, Schwinger, and Feynman," Phys. Rev. **75**, 486–502.

Frank, S., Badiey, M., Lynch, J., and Siegmann, W. (**2004**). "Analysis and modeling of broadband airgun data influenced by non-linear internal waves," J. Acoust. Soc. Am. **116**, 3404–3422.

Frank, S., Badiey, M., Lynch, J., and Siegmann, W. (**2005**). "Experimental evidence of three-dimensional acoustic propagation caused by nonlinear internal waves," J. Acoust. Soc. Am. **118**, 723–734.

Fredricks, A., Colosi, J. A., Lynch, J. F., Gawarkiewicz, G., Chiu, C. S., and Abbot, P. (**2005**). "Analysis of multipath scintillations observed during the summer 1996 New England shelfbreak PRIMER study," J. Acoust. Soc. Am. **117**, 1038–1057.

Headrick, R., Lynch, J., Kemp, J., Newhall, A., von der Heydt, K., Apel, J., Badiey, M., Chiu, C. S., Finette, S., Orr, M., Pasewark, B., Turgut, A., Wolf, S., and Tielguerger, D. (**2000a**). "Acoustic normal mode fluctuations statistics in the 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220.

Headrick, R., Lynch, J., Kemp, J., Newhall, A., von der Heydt, K., Apel, J., Badiey, M., Chiu, C. S., Finette, S., Orr, M., Pasewark, B., Turgut, A., Wolf, S., and Tielguerger, D. (**2000b**). "Modeling mode arrivals in the 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 221–236.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1993**). *Computational Ocean Acoustics* (Springer, New York).

Liu, A. (**1988**). "Analysis of nonlinear internal waves in the New York Bight," J. Geophys. Res. **93**, 12317–12329.

Liu, A., Ramp, S., Zhao, Y., and Tang, T. (**2004**). "A case study of internal solitary wave propagation during ASIAEX 2001," IEEE J. Ocean. Eng. **29**, 1144–1156.

Lynch, J. F., Colosi, J., Gawarkiewicz, G., Duda, T., Pierce, A., Badiey, M., Katsnelson, B., Miller, J., Seigmann, W., Chiu, C. S., and Newhall, A. (**2005**). "Consideration of finescale oceanography and 3-D acoustic effects for the ESME sound exposure model," IEEE J. Ocean. Eng. **31**, 33–48.

Munk, W. (**1974**). "Sound channel in an exponentially stratified ocean, with application to SOFAR," J. Acoust. Soc. Am. **55**, 220–226.

Preisig, J., and Duda, T. (**1997**). "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269.

Ramp, S., Yang, T., Duda, T., Lynch, J., Liu, A., Chiu, C. S., Bahr, F., Kim, H., and Yang, Y. (**2004**). "Internal solitons in the northeastern South china sea part 1: Sources and deep water propagation," IEEE J. Ocean. Eng. **29**, 1157–1181.

Rouseff, D., Turgut, A., Wolf, S., Finette, S., Orr, M., Pasewark, B., Apel, J., Badiey, M., Chiu, C. S., Headrick, R., Lynch, J., Kemp, J., Newhall, A., von der Heydt, K., and Tielburger, D. (**2002**). "Coherence of acoustic modes propagating through shallow water internal waves," J. Acoust. Soc. Am. **111**, 1655–1666.

Sakurai, J. J. (**1985**). *Modern Quantum Mechanics* (Addison-Westley, Redwood City, CA).

Tielburger, D., Finette, S., and Wolf, S. (**1997**). "Acoustic propagation through an internal wave field in a shallow water waveguide," J. Acoust. Soc. Am. **101**, 789–808.

Young, A. C. (**2006**). "Two dimensional acoustic propagation through oceanic internal solitary waves: Weak scattering theory and numerical simulation," Masters of Science Thesis, Department of Oceanography, Naval Postgraduate School, Monterey, CA.

Zhou, J., Zhang, X., and Rogers, D. (**1991**). "Resonant interaction of sound with internal solitons in a coastal zone," J. Acoust. Soc. Am. **90**, 2042–2054.

A method for approximating acoustic-field-amplitude uncertainty caused by environmental uncertainties

Kevin R. James and David R. Dowling[a)]
Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2133

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

In underwater acoustics, the accuracy of computational field predictions is commonly limited by uncertainty in environmental parameters. An approximate technique for determining the probability density function (PDF) of computed field amplitude, A, from known environmental uncertainties is presented here. The technique can be applied to several, N, uncertain parameters simultaneously, requires $N+1$ field calculations, and can be used with any acoustic field model. The technique implicitly assumes independent input parameters and is based on finding the optimum spatial shift between field calculations completed at two different values of each uncertain parameter. This shift information is used to convert uncertain-environmental-parameter distributions into PDF(A). The technique's accuracy is good when the shifted fields match well. Its accuracy is evaluated in range-independent underwater sound channels via an L_1 error-norm defined between approximate and numerically converged results for PDF(A). In 50-m- and 100-m-deep sound channels with 0.5% uncertainty in depth ($N=1$) at frequencies between 100 and 800 Hz, and for ranges from 1 to 8 km, 95% of the approximate field-amplitude distributions generated L_1 values less than 0.52 using only two field calculations. Obtaining comparable accuracy from traditional methods requires of order 10 field calculations and up to 10^N when $N>1$. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2950088]

PACS number(s): 43.30.Bp, 43.60.Cg, 43.60.Uv [RAS] Pages: 1465–1476

I. INTRODUCTION

The solutions to the partial differential equations governing many areas of science and engineering are often highly sensitive to changes in the boundary conditions. Thus, errors or uncertainties in boundary conditions produce corresponding errors and uncertainties in the predicted-field solutions that often overshadow all other sources of error. Since a direct relationship between boundary-condition parameters and predicted-field values is rarely analytical and invertible, standard error propagation techniques cannot be applied to determine predicted-field uncertainties. Thus, the task of determining predicted-field uncertainty from boundary-condition uncertainty must be application specific and must utilize the unique properties of the particular partial differential equations, boundary conditions, and field solutions of interest. The established approaches for this uncertainty-determination task, Monte Carlo and direct numerical sampling methods, typically involve a computational burden that increases exponentially with the number of uncertain parameters. This paper describes a technique to approximately determine the effect of environmental uncertainty on predicted-acoustic-field amplitudes in underwater sound channels that is more efficient.

It is important to make a distinction between uncertain environmental parameters, which produce uncertain field predictions, and randomly fluctuating environmental parameters, which produce random field solutions. The present work focuses only on the former, while the field of wave propagation in random media addresses the latter; a recent deep-water review can be found in Colosi *et al.* (1999). Uncertainty in shallow-ocean acoustics has received increased attention in recent years (see Livingston *et al.*, 2006 and references therein). Much of this research can be classified into three categories: determination of environmental parameter uncertainty, prediction of acoustic decision aid uncertainty, and prediction of acoustic field uncertainty. For example, in the first category, Gerstoft and Mecklenbräuker (1998) and Lin *et al.* (2006) address the problem of assessing uncertainty in environmental parameter inversion, when the input measurements occur in an uncertain environment. The technique described in this paper addresses the reverse question; what is the impact of environmental parameter uncertainty on the prediction of acoustic field amplitude?

In the second category, assessments of uncertainty in various sonar-based tactical decision aids for source detection and localization have also been completed. In Gerstoft *et al.* (2006), geoacoustic inversion data from the first category of research are used to estimate transmission loss in an uncertain field. A Monte Carlo approach to estimating the uncertainty of several relevant acoustic observables is outlined in Heaney and Cox (2006), and an application of Monte Carlo to matched-field processing can be found in Shorey *et al.* (1994). A maximum entropy method to calculate confidence intervals for detection of a high-frequency narrow-band source in an uncertain environment is given in Sibul *et al.* (2004). In addition, environmental uncertainty has been incorporated in a tracking and detection algorithm in order to obtain meaningful results even in the presence of large prediction errors (Stone and Osborn, 2004). Similarly, optimal sonar detection performance in diffuse noise has

a)Author to whom correspondence should be addressed.

been found to depend primarily on environmental uncertainties and the signal-to-noise ratio at the receivers (Sha and Nolte, 2005). The effect of water-depth mismatches on source localization via matched-field processing is explored in Shang and Wang (1991) and D'Spain *et al.* (1999). An improved source localization algorithm that provides more accurate results with imperfectly known surface statistics than previous methods is described in Haralabus *et al.* (1993). The effect of uncertainties in water depth, sound speed, and bottom characteristics on source localization uncertainties, using Bayesian inference theory, is examined in Dosso (2003) and Dosso *et al.* (2007a). Even if the field's distribution cannot be determined completely, upper or lower bounds for the uncertainty in various decision aids are highly useful in many applications. A method for obtaining a lower bound for the localization error in a sensor network is provided in Wang *et al.* (2004). In LePage and McDonald (2006), a lower limit is calculated for the deterioration of performance of time reversal in the presence of sound-speed uncertainties. A common feature of a large portion of the work referenced above is the reliance on either Monte Carlo or direct numerical sampling techniques to explicitly link environmental and acoustic-field uncertainties in the scenario(s) of interest.

In contrast, research efforts in the third category—where the present work belongs—focus on determining the uncertainty in acoustic field propagation calculations that arises from uncertainty in or imperfect knowledge of the acoustic environment without continuing on to assess how such uncertainties impact sonar performance. Here, the broadly applicable and mathematically sophisticated technique of polynomial chaos has been applied to sound propagation in uncertain ocean waveguides (Finette, 2005, 2006; Creamer, 2006). The polynomial chaos technique is able to provide uncertainty predictions throughout a spatial field for spatially varying uncertain parameters and its convergence criteria are best met when the uncertain spatially varying parameters have finite correlation length. The field-uncertainty prediction technique described in this paper differs from the polynomial chaos work in that it uses simpler mathematics, it provides uncertainty predictions at a selected point in space, and it addresses environmental parameters with infinite correlation length.

The goal of the present work is to produce an acoustic uncertainty-determination scheme that is simple (and fast) enough to be used in real time in sonar applications. To this end, the technique has only been developed for approximate determination of the probability density function (PDF) of the predicted-acoustic-field amplitude, A; phase uncertainty is not addressed. As shown herein, the technique may be readily applied to a countable number, N, of uncertain environmental parameters while only requiring $N+1$ field calculations. In addition, the technique is compatible with any type of acoustic propagation model, analytical or computational. The technique is found to work well when the changes in an uncertain parameter primarily lead to a spatial shift in the predicted acoustic field. Fortunately, such spatial field shifts can often be found (Dosso *et al.*, 2007b). In addition, optimum spatial shifting may provide the basis for a higher-moment model that closes acoustic-PDF transport equations (James and Dowling, 2005).

The remainder of the paper is divided into three sections. Section II describes the method when applied to one or more uncertain environmental parameters to produce a prediction for the resulting uncertainty in acoustic amplitude at a point in the field. Section III introduces an L_1 error-norm as a metric for evaluating the accuracy of approximate PDFs. Using this error-norm, the success of the approximate technique is reported for a variety of sound channels, of various dimensions, bottom properties, water-column properties, and source frequencies. The accuracy of the proposed method is then compared to a simple linear approximation technique that utilizes the same number of field calculations and to the more-computationally-intense direct numerical sampling method. Section IV summarizes the findings, including the successes and limitations of the approximate technique, and also states the conclusions drawn from this study.

II. ACOUSTIC AMPLITUDE PDF PREDICTION FROM OPTIMUM SPATIAL SHIFTS

This section describes an approximate technique for determining the PDF of acoustic amplitude for N uncertain environmental parameters based on $N+1$ acoustic field calculations. The inputs are an acoustic-field calculation routine, the environmental parameters necessary for completing a field calculation, and the PDFs for the uncertain environmental parameters. First, the technique is illustrated via an example calculation conducted at the range-depth location (r,z) for a single uncertain parameter. Instructions and formulae for extending the technique to multiple uncertain parameters follow.

For one uncertain environmental parameter, an approximate field-amplitude PDF at any point can be computed using two range-depth acoustic-field calculations. The examples shown here are from range-independent sound channels and rely on KRAKEN (Porter and Reiss, 1984) for the field calculations with a unity strength source. The basic idealized range-independent sound channel used in this study is shown in Fig. 1. It is specified by eight parameters: overall sound channel depth H, mixed-layer depth d_1, mixed-layer sound speed c_1, thermocline lower limit d_2, deep water-column sound speed c_2, bottom sound speed c_3, bottom density ρ_3, and bottom absorption coefficient γ_3. Of course more sophisticated sound channel parametrizations that include multiple bottom layers, for example, are possible but were not pursued here. For the current example, the water-column depth, H, is presumed uncertain but the distribution of possible H-values, PDF(H), is presumed known. The first step is to perform two field calculations at differing values of the uncertain parameter. Here, this means one field calculation with $H=\langle H\rangle$ (the reference calculation) and a second with $H=\bar{H}$ (the sensitivity-assessment calculation). Setting $\bar{H}=\langle H\rangle+\sigma_H$ is recommended, where $\langle H\rangle$ and σ_H are the expected value and standard deviation of H, respectively.

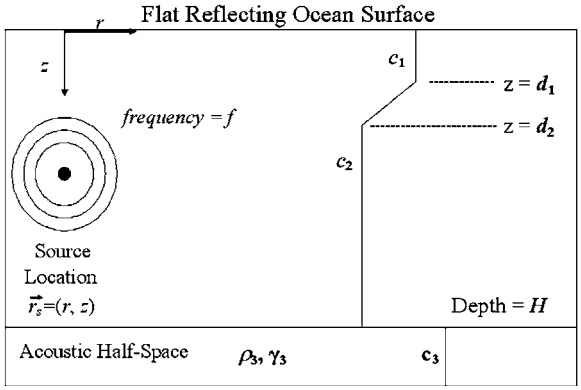


FIG. 1. Schematic of a generic range-independent sound channel that is described by eight parameters: overall depth H, mixed-layer depth d_1, thermocline lower limit d_2, mixed-layer sound speed c_1, deep water-column sound speed c_2, bottom sound speed c_3, bottom density ρ_3, and bottom absorption coefficient γ_3. Here mean water-column depths of 50 and 100 m are considered along with bottom types typical of sand, silt, and gravel.

Throughout this manuscript uncertain input parameters are assumed to be Gaussian distributed, reference and sensitivity-assessment calculations are taken one standard deviation apart, $\langle\ \rangle$-brackets indicate an expected value, and $\sigma_{(\ )}$ is the standard deviation of ( ). In the present example with uncertain H, the lower portion of the sound channel is merely extended for the sensitivity-assessment calculation; the channel was not uniformly stretched.

The second step involves selecting the range, R, and depth, Z, dimensions of windows in the computed fields that will be used to determine the optimal spatial shift. Smaller windows allow for better local approximations; however, the window must also be large enough to contain the approximate spatial shift corresponding to the $\bar{H}-\langle H\rangle$ change in uncertain input. It is possible to optimize the window-size choice via an iterative scheme but results from such an approach are not reported here. In the present effort, the following formulas were used:

$$Z = 2.0c_2/f, \tag{1a}$$

and

$$R = 0.15Z(r/\langle H\rangle)\tan\,\theta_c, \tag{1b}$$

where f is the acoustic frequency (in hertz) and $\theta_c = \sin^{-1}(c_2/c_3)$ is the critical angle for penetration into the bottom of the idealized sound channel. The approximate PDF results were only mildly sensitive to the constants, 2.0 and 0.15, in the window-size formula; changing them by $\pm 30\%$ only caused mild degradation of the approximate field-shifting technique's accuracy.

The third step involves finding the optimum spatial shifts, Δr_o and Δz_o, that minimize the normalized root-mean-square amplitude difference, D, between the two windowed sound fields:

$$D(r,z,\langle H\rangle;\Delta r,\Delta z,\bar{H}-\langle H\rangle) = \left[\frac{1}{(Z-\Delta z)(R-\Delta r)} \times \int_{r+\Delta r-R/2}^{r+R/2}\int_{z+\Delta z-Z/2}^{z+Z/2}\left|\frac{A(r'-\Delta r,z'-\Delta z,\bar{H})}{\hat{A}(r-\Delta r,z-\Delta z,\Delta r,\Delta z,\bar{H})} - \frac{A(r',z',\langle H\rangle)}{\hat{A}(r,z,\Delta r,\Delta z,\langle H\rangle)}\right|^2 dr'dz'\right]^{1/2}, \tag{2a}$$

where

$$\hat{A}(r,z,\Delta r,\Delta z,H) = \frac{1}{(Z-\Delta z)(R-\Delta r)} \times \int_{r+\Delta r-R/2}^{r+R/2}\int_{z+\Delta z-Z/2}^{z+Z/2} A(r',z',H)dr'dz'. \tag{2b}$$

$A(r,z,H)$ is the field amplitude at (r,z) when the water-column depth is H and the integration takes place in the region of the overlap between the two field windows. Here, $\hat{A}(r,z,\Delta r,\Delta z,H)$ is the average amplitude in the overlap region for the field calculated at water-column depth H. The limits of integration and prefactors shown in Eqs. (2a) and (2b) presume Δr and Δz to both be positive. When Δr is negative the upper and lower range-integration limits change to $r+\Delta r+R/2$ and $r-R/2$, respectively, and the range-integration prefactor becomes $1/(R+\Delta r)$. Similarly, when Δz is negative the upper and lower depth-integration limits change to $z+\Delta z+Z/2$ and $z-Z/2$, respectively, and the depth-integration prefactor becomes $1/(Z+\Delta z)$. Since the field amplitudes in each shifted window are divided by the mean amplitude in the window, D is a dimensionless measure of the accuracy of each possible shift. Figure 2 shows contour plots of two amplitude fields at $f=400$ Hz centered on $(r,z)=(3.0\text{ km},25\text{ m})$ when $\bar{H}-\langle H\rangle=\sigma_H=0.25$ m, $H=\langle H\rangle=50$ m, $c_1=\langle c_1\rangle=1500$ m/s, $d_1=\langle d_1\rangle=10$ m, $c_2=\langle c_2\rangle=1450$ m/s, $d_2=\langle d_2\rangle=20$ m, $c_3=\langle c_3\rangle=1575$ m/s, $\rho_3=\langle\rho_3\rangle=1700$ kg/m^3, and $\gamma_3=\langle\gamma_3\rangle=1.0$ dB/wavelength. Nominal bottom properties throughout this manuscript are taken from Table 1.3 in Jensen *et al.* (1994). Here the optimum shifts are $\Delta r_o=26.12$ m in range and $\Delta z_o=0.13$ m in depth. These optimum spatial shifts approximately quantify the effect of changing H from $\langle H\rangle$ to $\bar{H}$.

The fourth step involves extending the approximate-shift relationship to larger or smaller changes in the uncertain parameter. To accomplish this, the optimum spatial shifts are used to determine constants in assumed proportionalities between changes in uncertain parameters and shifts in the spatial coordinates. For the present example involving uncertain channel depth, H, these proportionalities are

$$\frac{\Delta r}{r} = \alpha\frac{(H-\langle H\rangle)}{\langle H\rangle}, \tag{3a}$$

and

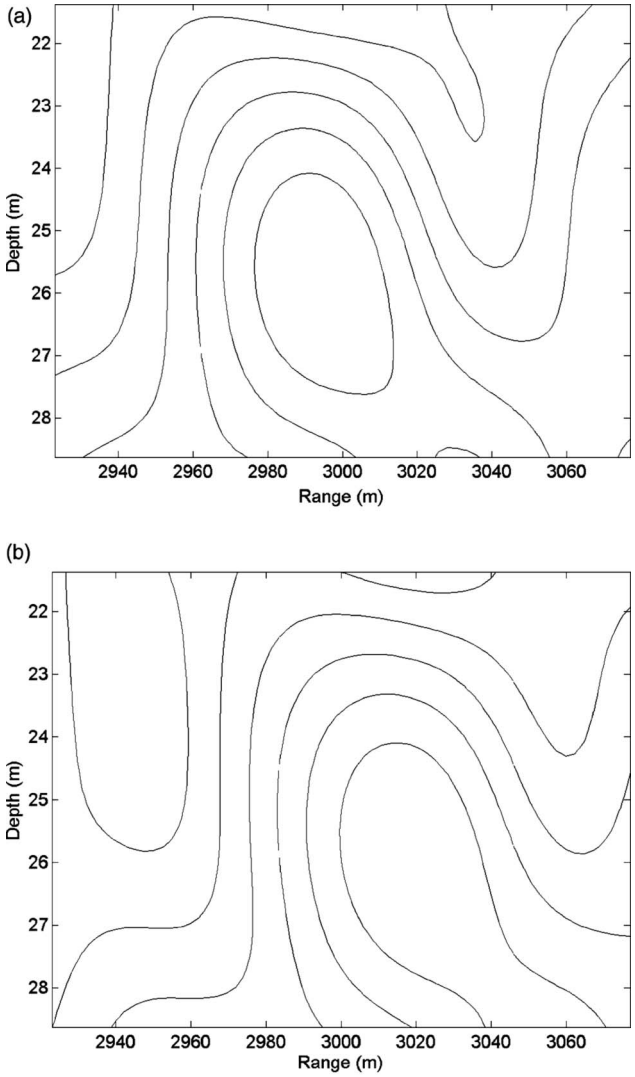


FIG. 2. Representative contour plots of small sections of an acoustic field at $f=400$ Hz centered on $(r,z)=(3.0\ \text{km},25\ \text{m})$ in a sound channel when the channel depth is $\langle H\rangle=50$ m (a) and when it is $\langle H\rangle+\sigma_H=50.25$ m (b).

$$\frac{\Delta z}{z}=\beta\frac{(H-\langle H\rangle)}{\langle H\rangle}. \tag{3b}$$

The proportionality constants α and β are determined from the optimum shift information. These relationships are inspired by the form of the waveguide invariant (Brekhovskikh and Lysanov, 1991, D'Spain and Kuperman, 1999), and the observation that small mismatches or changes in environmental parameters often lead to spatial shifts in matched field source location (D'Spain *et al.*, 1999), and in measured and computed acoustic fields (Kim *et al.*, 2003, Dosso *et al.*, 2007b). Naturally, Eqs. (3a) and (3b) provide just one simple parametrization; other choices are possible. Thus, the calculations for the fourth step merely involve inverting Eqs. (3a) and (3b) with $H=\bar{H}$ to find the constants:

$$\alpha=\left(\frac{\Delta r_o}{r}\right)\Bigg/\left(\frac{\bar{H}-\langle H\rangle}{\langle H\rangle}\right), \tag{4a}$$

and

$$\beta=\left(\frac{\Delta z_o}{z}\right)\Bigg/\left(\frac{\bar{H}-\langle H\rangle}{\langle H\rangle}\right). \tag{4b}$$

For the example fields shown in Fig. 2, $\alpha=1.74$ and $\beta=1.04$. With these two constants determined, the field amplitude at the point of interest (r,z) for an arbitrary-water-column depth, H, can be approximated by the field amplitude at a shifted location in the field calculated at $H=\langle H\rangle$. Since the acoustic field was also calculated at $H=\bar{H}$, the arbitrary-water-column depth field amplitude can also be approximated via spatial shifting in this second field calculation. These shifts are of the same form as Eqs. (3a) and (3b):

$$\frac{\Delta\bar{r}}{r}=\bar{\alpha}\frac{(H-\bar{H})}{\bar{H}}, \tag{5a}$$

and

$$\frac{\Delta\bar{z}}{z}=\bar{\beta}\frac{(H-\bar{H})}{\bar{H}}, \tag{5b}$$

where $\bar{\alpha}$ and $\bar{\beta}$ are found via algebraic inversions similar to Eqs. (4a) and (4b),

$$\bar{\alpha}=\left(\frac{-\Delta r_o}{r}\right)\Bigg/\left(\frac{\langle H\rangle-\bar{H}}{\bar{H}}\right), \tag{6a}$$

and

$$\bar{\beta}=\left(\frac{-\Delta z_o}{z}\right)\Bigg/\left(\frac{\langle H\rangle-\bar{H}}{\bar{H}}\right). \tag{6b}$$

Linear interpolation between the two approximations yields a final weighted approximation for $A(r,z,H)$,

$$A(r,z,H)\approx A(r-\Delta r,z-\Delta z,\langle H\rangle)+\left(\frac{H-\langle H\rangle}{\bar{H}-\langle H\rangle}\right)[A(r-\Delta\bar{r},z-\Delta\bar{z},\bar{H})-A(r-\Delta r,z-\Delta z,\langle H\rangle)]. \tag{7}$$

Equation (7) provides the fundamental simplification of the approximate field-shifting technique. The advantage lies in the fact that $A(r,z,\langle H\rangle)$ and $A(r,z,\bar{H})$ have been calculated in the first step while $A(r,z,H)$ has not. Thus, Eq. (7) allows many field calculations at different H-values to be bypassed. However, it is approximate as shown in Fig. 3(a) where amplitude versus water-column depth sensitivity curves are plotted at a receiver range of 6.0 km and depth of 25 m. The two curves in Fig. 3(a) are similar but not identical. The dashed curve from Eq. (7) is based on two field calculations while the solid curve was obtained from 41 field calculations at H-values spanning $\langle H\rangle\pm3\sigma_H$. The straight dotted line in Fig. 3(a) is a simple two-point linear fit to the values of $A(r,z,\langle H\rangle)$ and $A(r,z,\bar{H})$ at the spatial location of interest:

$$A(r,z,H)\approx A(r,z,\langle H\rangle)+\left(\frac{A(r,z,\bar{H})-A(r,z,\langle H\rangle)}{\bar{H}-\langle H\rangle}\right)(H-\langle H\rangle). \tag{8}$$

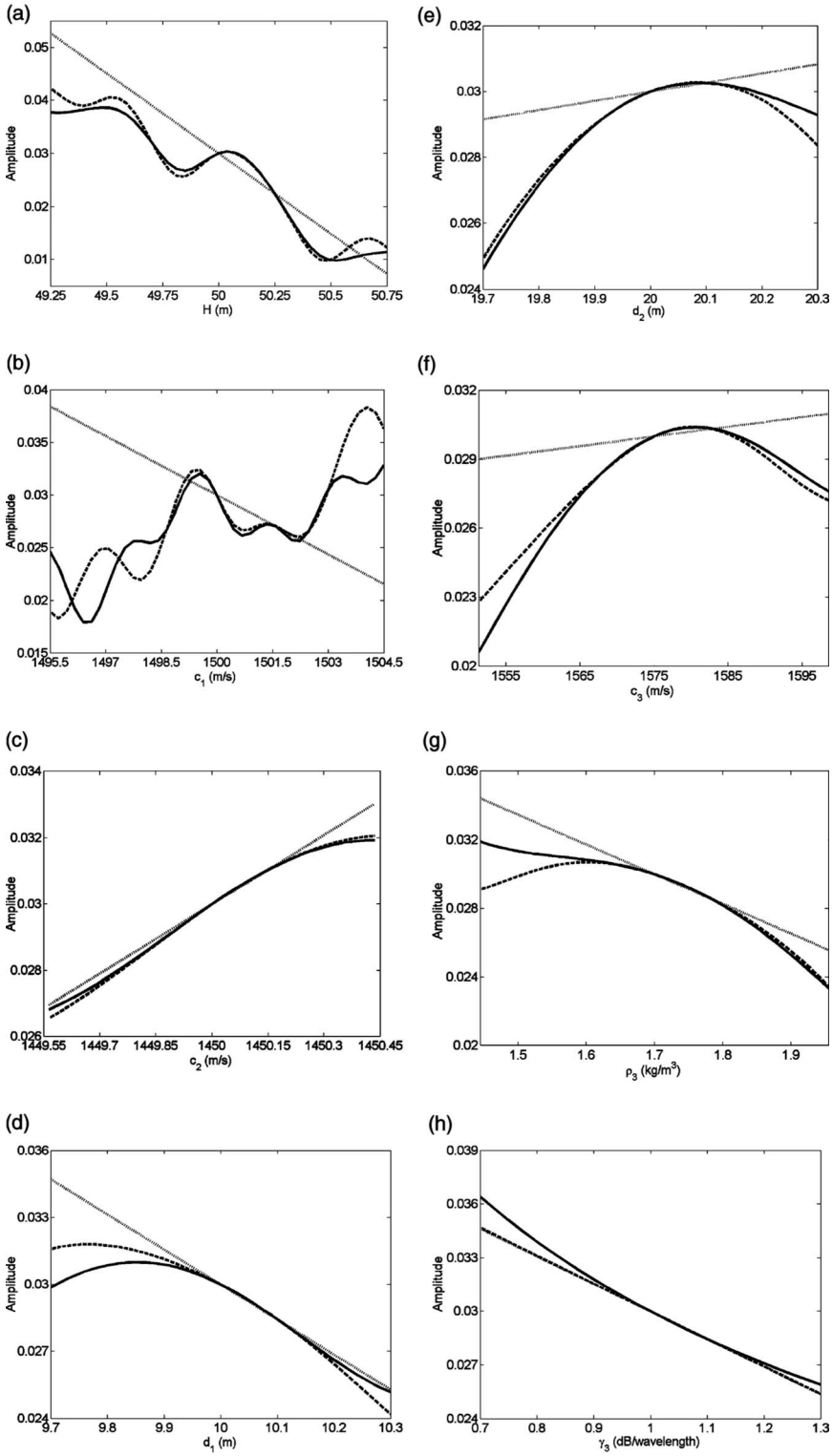


FIG. 3. Sensitivity curve comparisons at $f=400$ Hz and $(r,z)=(6.0\ \text{km}, 25\ \text{m})$ between approximate results from Eqs. (7) (dashed curve) and (8) (dotted line) using field computations similar to that shown in Fig. 2 and numerically converged results (solid curve) based on 41 field computations for each of the eight parameters that specify the Fig. 1 sound channel: (a) uncertain depth with $\langle H\rangle=50$ m and $\sigma_H/\langle H\rangle=0.005$, (b) uncertain mixed-layer sound speed with $\langle c_1\rangle=1500$ m/s and $\sigma_{c_1}/\langle c_1\rangle=0.001$, (c) uncertain deep-channel sound speed with $\langle c_2\rangle=1450$ m/s and $\sigma_{c_2}/\langle c_2\rangle=0.0001$, (d) uncertain mixed-layer depth with $\langle d_1\rangle=10$ m and $\sigma_{d_1}/\langle d_1\rangle=0.01$, (e) uncertain depth to the bottom of the thermocline with $\langle d_2\rangle=20$ m and $\sigma_{d_2}/\langle d_2\rangle=0.005$, (f) uncertain bottom sound speed with $\langle c_3\rangle=1575$ m/s and $\sigma_{c_3}/\langle c_3\rangle=0.005$, (g) uncertain bottom density with $\langle \rho_3\rangle=1700\ \text{kg/m}^3$ and $\sigma_{\rho_3}/\langle \rho_3\rangle=0.05$, and (h) uncertain bottom sound absorption coefficient with $\langle \gamma_3\rangle=1.0$ dB/ wavelength and $\sigma_{\gamma_3}/\langle \gamma_3\rangle=0.10$ (note that the dashed and dotted curves overlap for this case). In each plot, the horizontal axis spans three standard deviations of the uncertain parameter.

The remaining frames of Fig. 3 show approximate field-shift sensitivity curves obtained from Eq. (7) when the various environmental parameters of the Fig. 1 sound channel are individually uncertain. The line types in Fig. 3(a) are followed in the subsequent frames of Fig. 3. These calculations were performed for the same frequency, geometry, and environmental parameters as Fig. 2. The standard deviations of the various input parameters are listed in the Fig. 3 caption. In this case, the various results show good agreement between the results of Eq. (7), when applied to each environmental parameter, and direct calculations, even though the various sensitivity curves are clearly nonlinear.

The fifth and final step of the technique is to generate an approximate field-amplitude PDF from the sensitivity data using standard transformation techniques (see Papoulis, 1965) and the known input uncertain-parameter distribution, PDF(H) in this example. Figure 4(a) shows both the approximate and numerically converged PDFs obtained from the sensitivity curves in Fig. 3(a), when the input water-column-depth distribution has $\sigma_H=0.25$ m. Note that the amplitude

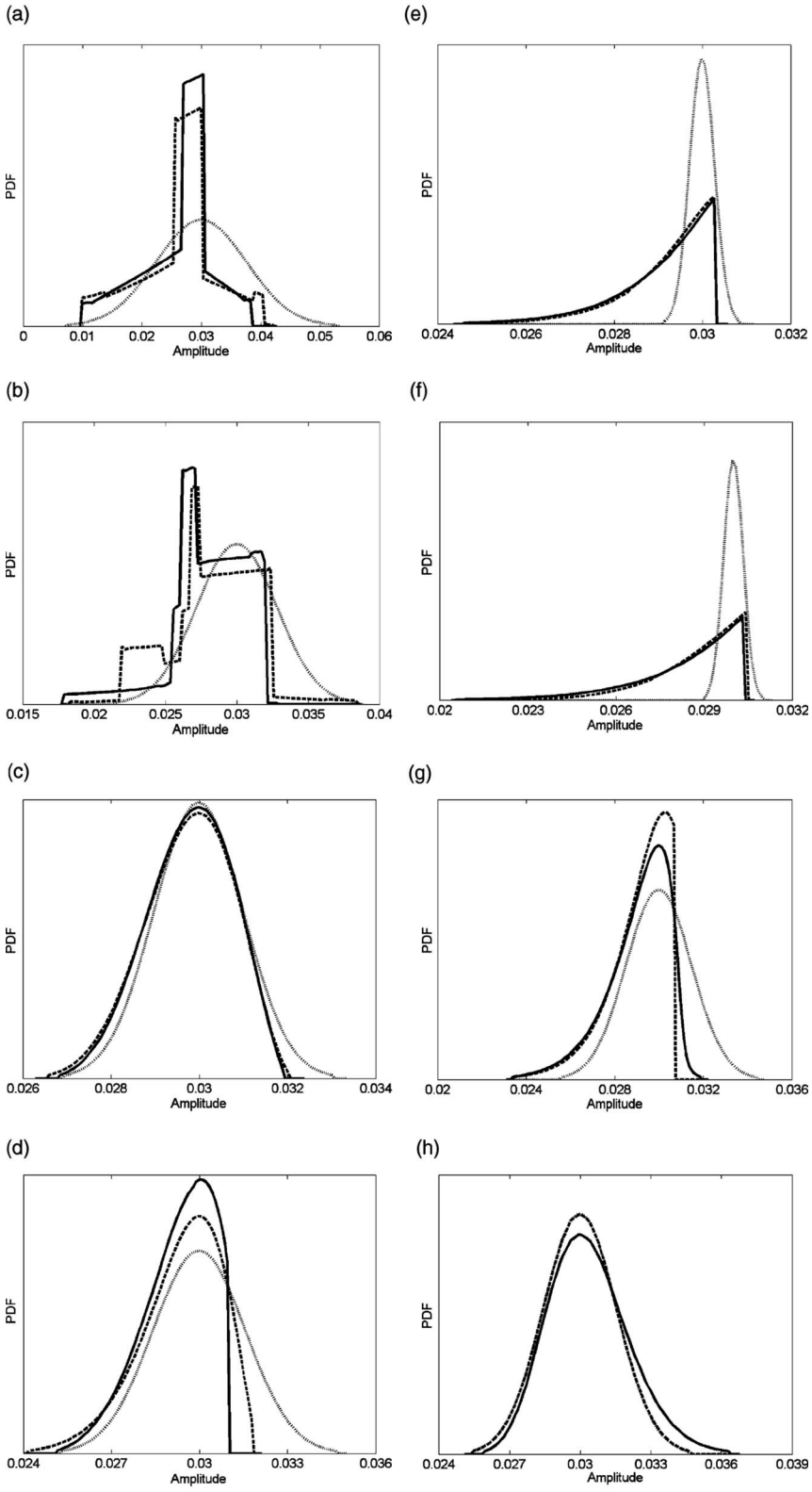


FIG. 4. Same as Fig. 3, except here the acoustic-field-amplitude PDFs are plotted for the case when each of the eight parameters that specify the Fig. 1 sound channel is uncertain.

PDF is approximated well by the technique described here. Furthermore, the actual amplitude PDF is clearly non-Gaussian even though the uncertainty is only 0.5% of the mean depth, and the input PDF is Gaussian. This happens because the acoustic amplitude depends on water depth in a nonlinear manner, and, in this case, values of H above and below $\langle H \rangle$ can produce acoustic amplitudes that are below that produced by $H \approx \langle H \rangle$. The remaining frames of Fig. 4 show PDF(A) obtained from numerically converged direct-sampling calculations, the field-shifting technique, and simple linear fitting when the various environmental parameters of the Fig. 1 sound channel are individually uncertain. In Fig. 4 the line types, frequency, geometry, and environmental parameters and their uncertainties are the same as those in Fig. 3. In each case, the field-shift-produced PDF(A) show better agreement with the numerically converged

PDF(A) than the PDF(A) developed from simple linear fitting. However, the two approximate approaches will perform similarly when the relevant amplitude sensitivity curve is linear, such as in Figs. 3(h) and 4(h).

The field-shifting technique can be readily extended to multiple uncertain input parameters with a few modifications to the five steps enumerated above. Here, for simplicity, independence of the input uncertain parameters is assumed as well. Incorporation of dependencies between uncertain inputs, while not inconsistent with the proposed approximation method, was beyond the scope of this effort.

For multiple uncertain parameters, the first four steps of the single-parameter method are performed independently for each uncertain input parameter, ψ_i, $1 \le i \le N$. This involves one reference calculation and N sensitivity-assessment field calculations. The N-uncertain-parameter extension of Eqs. (3a) and (3b) can be readily formulated:

$$\frac{\Delta r}{r} = \sum_{i=1}^{N} \alpha_i \frac{(\psi_i - \langle \psi_i \rangle)}{\langle \psi_i \rangle}, \tag{9a}$$

and

$$\frac{\Delta z}{z} = \sum_{i=1}^{N} \beta_i \frac{(\psi_i - \langle \psi_i \rangle)}{\langle \psi_i \rangle}, \tag{9b}$$

where the α_i and β_i are determined from

$$\alpha_i = \left(\frac{(\Delta r_o)_i}{r} \right) \Bigg/ \left(\frac{\bar{\psi}_i - \langle \psi_i \rangle}{\langle \psi_i \rangle} \right), \tag{10a}$$

and

$$\beta_i = \left(\frac{(\Delta z_o)_i}{z} \right) \Bigg/ \left(\frac{\bar{\psi}_i - \langle \psi_i \rangle}{\langle \psi_i \rangle} \right), \tag{10b}$$

where $(\Delta r_o)_i$ and $(\Delta z_o)_i$ are the optimum range and depth shifts for variations in ψ_i. The N-uncertain-parameter extensions of Eqs. (5a), (5b), (6a), and (6b) follow from Eqs. (9a), (9b), (10a), and (10b). The N-uncertain-parameter extension of Eq. (7) is

$$A(r, z, \psi_1, \psi_2, \ldots, \psi_N) \approx A_0 + \sum_{i=1}^{N} \left(\frac{\psi_i - \langle \psi_i \rangle}{\bar{\psi}_i - \langle \psi_i \rangle} \right) [A_i - A_0], \tag{11}$$

where $A_0 = A(r - \Delta r, z - \Delta z, \langle \psi_1 \rangle, \langle \psi_2 \rangle, \ldots, \langle \psi_N \rangle)$ is the reference field calculation evaluated at the appropriate-shift location, and the $A_i = A(r - (\Delta \bar{r})_i, z - (\Delta \bar{z})_i, \langle \psi_1 \rangle, \langle \psi_2 \rangle, \ldots, \langle \psi_{i-1} \rangle, \bar{\psi}_i, \langle \psi_{i+1} \rangle, \ldots, \langle \psi_N \rangle)$ are the N sensitivity-assessment calculations, each evaluated at its appropriate-shift location. Similarly, the N-uncertain-parameter extension of Eq. (8), the simple linear approximation, is

$$\begin{aligned} A(r, z, \psi_1, \psi_2, \ldots, \psi_N) &= A(r, z, \langle \psi_1 \rangle, \langle \psi_2 \rangle, \ldots, \langle \psi_N \rangle) + \sum_{i=1}^{N} \left(\frac{\psi_i - \langle \psi_i \rangle}{\bar{\psi}_i - \langle \psi_i \rangle} \right) \\ &\times [A(r, z, \langle \psi_1 \rangle, \langle \psi_2 \rangle, \ldots, \langle \psi_{i-1} \rangle, \bar{\psi}_i, \langle \psi_{i+1} \rangle, \ldots, \langle \psi_N \rangle) \\ &- A(r, z, \langle \psi_1 \rangle, \langle \psi_2 \rangle, \ldots, \langle \psi_N \rangle)]. \end{aligned} \tag{12}$$

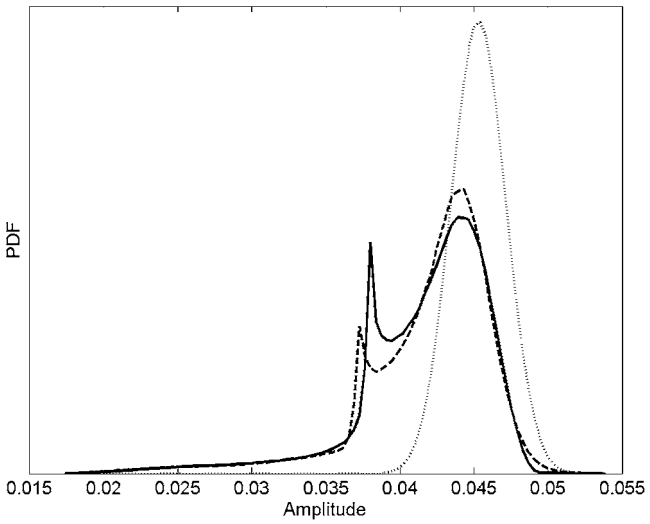


FIG. 5. Approximate and numerically converged PDFs of the acoustic field amplitude for two uncertain parameters: depth, H, and mixed-layer sound speed, c_1, at $f=250$ Hz and $r=4.0$ km with all other parameters, including line types, the same as for Fig. 3. The numerically converged PDF comes from more than 400 field calculations.

To illustrate these formulas, consider the Fig. 1 sound channel with uncertain depth H and mixed-layer sound speed c_1 at $f=250$ Hz and $r=4.0$ km, with all other parameters evaluated at the mean values given in Fig. 3. Here, three field calculations were performed for the field shifting and simple linear techniques—the reference calculation at $\langle H \rangle$ and $\langle c_1 \rangle$, the depth sensitivity-assessment calculation at $\bar{H} = \langle H \rangle + \sigma_H$ and $\langle c_1 \rangle$, and the mixed-layer sound-speed sensitivity-assessment calculation at $\langle H \rangle$ and $\bar{c}_1 = \langle c_1 \rangle + \sigma_{c_1}$. Figure 5 shows the amplitude PDF results from the approximate field-shifting technique (dashed curve), numerically converged direct sampling (solid curve), and simple linear fitting (dotted curve) for $\sigma_H / \langle H \rangle = 0.005$ and $\sigma_{c_1} / \langle c_1 \rangle = 0.001$. In this case, the field-shifting PDF is clearly more accurate than the one obtained from simple linear fitting. For reference, the direct-sampling PDF shown in Fig. 5 required more than 400 field calculations.

When multiple parameters are uncertain, the proportionality constants, α_i and β_i, may depend slightly on the values of the other uncertain parameters, ψ_j for $j \neq i$. However, anticipated applications of the approximate field-shifting technique are likely to involve only small relative changes in the uncertain parameters, so any dependence of the proportionality constants on the other environmental parameters is suppressed. Although this suppression reduces accuracy, this accuracy reduction primarily occurs in cases where the uncertain input parameters have relatively wide distributions and the utility of acoustic field predictions may have been lost because the final amplitude uncertainties are many decibels.

A comparison of numerically converged and approximate PDFs is shown in Fig. 6 when three environmental parameters are uncertain. The line types and parameters are the same as in Figs. 3–5, except that for this example $f = 600$ Hz and $r = 1.0$ km with $\sigma_{d_1} / \langle d_1 \rangle = 0.01$. Again the field-shifting PDF is closer to the direct-sampling result than the

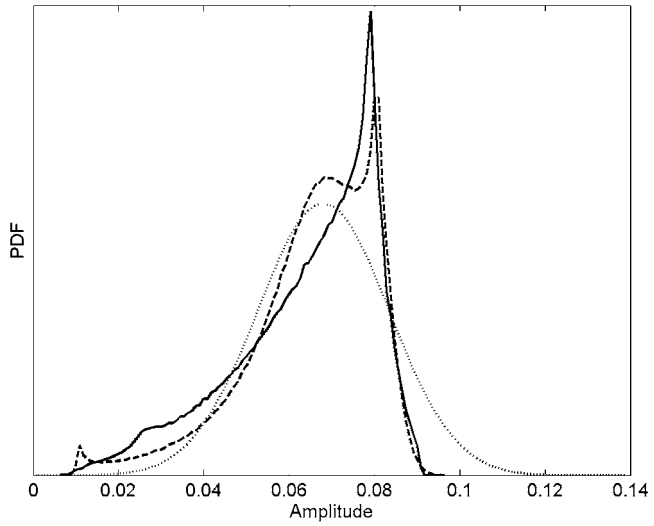


FIG. 6. Approximate and numerically converged PDFs of the acoustic field amplitude for three uncertain parameters: depth, H, mixed-layer sound speed, c_1, and mixed-layer depth, d_1, at $f=600$ Hz and $r=1.0$ km with all other parameters, including line types, the same as for Fig. 3. The numerically converged PDF comes from more than 9000 field calculations.

one obtained from simple linear fitting. Here, the direct-sampling PDF shown in Fig. 6 required more than 9000 field calculations.

As might be expected, when uncertainties in the input parameters and the number of uncertain input parameters both increase, larger discrepancies are possible between the approximate and numerically converged distributions. In practice, however, this loss of accuracy is balanced by the fact that as the input parameters, and thus the acoustic amplitude, become more uncertain, precise information about the amplitude distribution becomes less important for the anticipated applications of this technique. For example, a prediction of 12 dB of amplitude uncertainty may have the same impact on a sonar decision aid as 10 or 15 dB of uncertainty.

III. ASSESSMENT OF ACCURACY

The examples in the previous section demonstrate that it is possible to apply the approximate-PDF technique with great computational savings, but they do not provide a statistically compelling indication of the reliability or the accuracy of the technique. In this section, the accuracy of this technique is explored over variations in sound channel parameters. This technique's accuracy is then compared to simple linear PDF transformation using the same number, $N+1$, of field calculations and to the direct numerical sampling method, which has a computational burden that increases exponentially with N.

Although the accuracy of an approximate PDF can be reported by comparing its moments to those determined from an exact distribution, approximate PDF accuracy is quantified here via an L_1 error-norm:

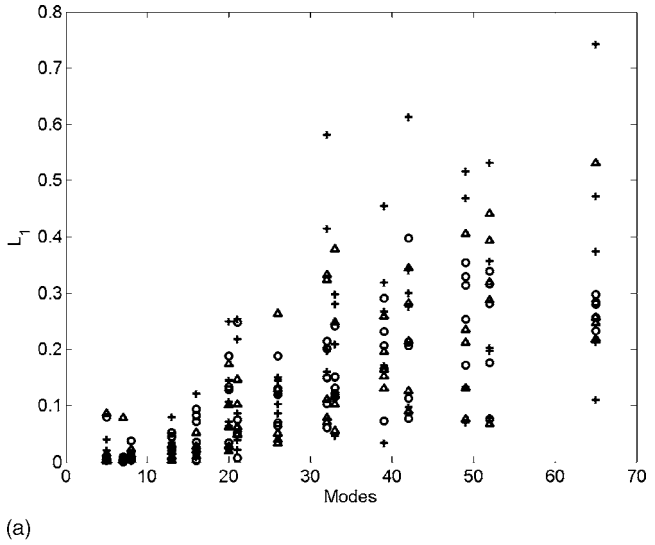
$$L_1 = \int_o^\infty |\mathrm{PDF}_a(A) - \mathrm{PDF}_c(A)|dA, \tag{13}$$

where PDF_a and PDF_c are approximate and numerically converged PDFs, respectively. This choice was made because L_1 provides a dimensionless single-number overall evaluation of PDF matching and because it has a simple geometric interpretation as the nonoverlapping area of PDF_a and PDF_c. in plots like Figs. 4–6. The more common L_2 error-norm, which would have $|\mathrm{PDF}_a(A)-\mathrm{PDF}_c(A)|^2$ as the integrand, was bypassed for accuracy assessments because it is not dimensionless. The lowest possible value for L_1 is zero, indicating a perfect match between the two PDFs (perfect accuracy), while the highest possible value for the L_1 is 2, indicating no overlap between the two PDFs (complete inaccuracy).

For the L_1-calculations, the numerically converged PDF in Eq. (13) was obtained from direct numerical sampling. For each uncertain parameter—for example, H—direct numerical sampling involves K equally-spaced-in-H computations of $A(r,z,H)$ that span the range of possible H-values. For N uncertain parameters, this corresponds to K^N field calculations. Interpolation in the resulting data yields an acoustic-amplitude sensitivity curve (or surface) for changes in all uncertain inputs. These sensitivity data, combined with the distributions for all the uncertain inputs, yield a numerically converged amplitude PDF via standard PDF transformation techniques, as noted for the field-shifting technique. The value of K required to generate smooth $A(r,z,H)$ and $\mathrm{PDF}(A)$ curves varies with the complexity of the sensitivity curves; values between 11 and 41 were used for the calculations presented here.

The L_1 values found in this investigation varied from less than 0.001 to more than 1.0. For the PDFs shown in Figs. 4(a)–4(h), 5, and 6, the L_1 values are 0.26, 0.42, 0.03, 0.23, 0.06, 0.15, 0.16, .015, 0.13, and 0.20, respectively, for the field-shifting-technique PDFs and are 0.71, 0.52, 0.12, 0.52, 1.06, 1.25, 0.49, 0.16, 0.88, and 0.37, respectively, for the simple-linear-fit PDFs. Furthermore, the means and standard deviations of the field-shifting-technique PDFs are within 2% and 20%, respectively, of the numerically converged values. Such L_1 and moment-difference errors are presumed to be small enough for successful use of the approximate field-shifting PDF technique in many applications for uncertainty analysis.

The overall accuracy of the approximate-PDF technique for a single uncertain parameter—the sound channel depth H—in the idealized Fig. 1 sound channel was assessed by performing many calculations at several receiver locations $[r=1,2,4,6,$ 8 km, $z=(\langle d_1\rangle+\langle d_2\rangle)/2$, $\langle H\rangle/2$, and $3\langle H\rangle/4]$ for a variety of environmental parameters to obtain a broad sample of L_1 values. The source was always located at $r=0$ and $z=\langle H\rangle/2$, and the input channel-depth distribution, $\mathrm{PDF}(H)$, had standard deviation of $\sigma_H/\langle H\rangle=0.005$. This input uncertainty level was chosen to incorporate cases of high and low amplitude uncertainties as well as high and low approximation accuracies. The results of this study are shown in Fig. 7 which provides two scatter plots of L_1 versus the number of propagating modes for sound frequencies be-



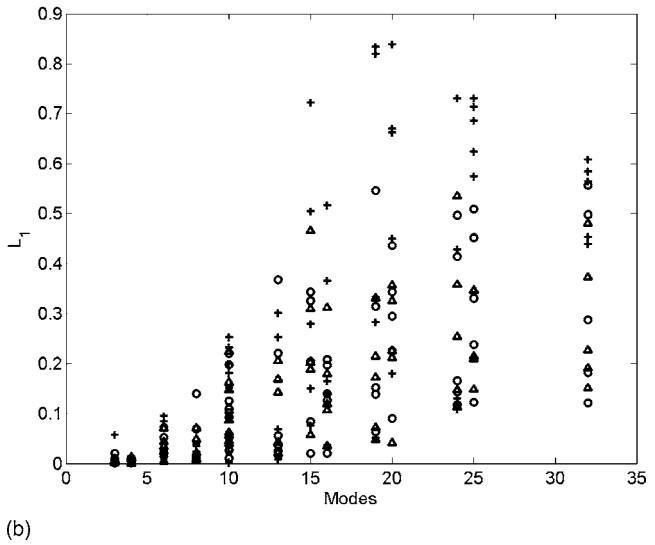


FIG. 7. Scatter plots of L_1-values vs the number of propagating modes in the sound channel for frequencies from 100 to 800 Hz and source receiver ranges from 1 to 8 km: (a) 100-m-deep channel with receiver depths of 15 m (pluses), 50 m (circles), and 75 m (triangles) and (b) 50-m-deep channel with receiver depths of 15 m (pluses), 25 m (circles), and 37.5 m (triangles). All other parameters are as specified in Fig. 3.

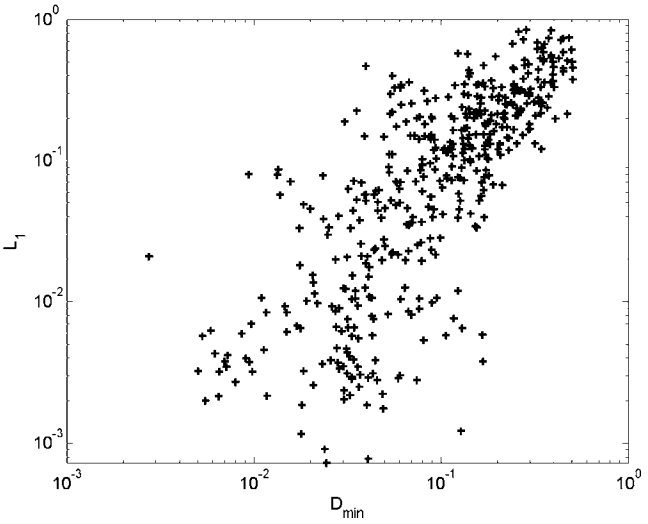


FIG. 8. Log-log plot of L_1 vs the normalized root-mean-square amplitude difference at the optimum field shift, $D_{\min}$, for the calculations shown in Fig. 7. Although there is scatter, increasing L_1 is clearly correlated with increasing $D_{\min}$.

tween 100 and 800 Hz and bottom properties of soft silt (c_3=1575 m/s, ρ_3=1700 kg/m^3, γ_3=1.0 dB/wavelength), sand (c_3=1650 m/s, ρ_3=1900 kg/m^3, γ_3=0.8 dB/wavelength), and gravel (c_3=1800 m/s, ρ_3=2000 kg/m^3, γ_3=0.6 dB/wavelength) when $\langle H\rangle$=100 m [Fig. 7(a)] and when $\langle H\rangle$=50 m [Fig. 7(b)]. The other water-column properties were not altered from their Fig. 2 values.

The primary finding is that the number of modes in the sound channel and the quality of initial shifted-field matching (the third step in the technique's basic procedure) were the best indicators for the potentially high L_1 values for a single uncertain variable. Here, the number of propagating modes was determined directly from the KRAKEN output and it is used as a surrogate for the complexity of acoustic propagation in the idealized sound channels. Comparisons of the panels of Fig. 7 show that L_1 increases when the sound channel is shallower and the sound-speed's depth dependence occupies a larger fraction of the channel depth. However, the overall assessment of the approximate field-shifting technique is positive, as 95% of all the cases investigated with a 0.5% uncertainty in channel-depth return L_1 values less than 0.52 even though 26% of the points in Fig. 7 have $\sigma_A/\langle A\rangle$ greater than 0.4.

The outlying high-L_1 points occur when the difference between $A(r,z,\langle H\rangle)$ and $A(r,z,\bar{H})$ in the local region around (r,z) is not readily described by a spatial shift. As Fig. 7 illustrates, this can occur as the number of propagating modes in the field increases, but for an arbitrary uncertain input parameter there may exist conditions for which small changes do not correspond to spatial shifts. However, the occurrence of high-L_1 values can be anticipated without knowing PDF$_c(A)$. Figure 8 shows a log-log scatter plot of all the Fig. 7 L_1 data versus $D_{\min}=D(r,z,\langle H\rangle;\Delta r_o,\Delta z_o,\sigma_h)$. Although the L_1 points are scattered, the trend of increasing L_1 with increasing $D_{\min}$ is clear; when $D_{\min}$ is less than 0.25, 95% of the approximate field-shift-produced PDFs generated L_1 values below 0.35. Higher values of $D_{\min}$ indicate that changes in the uncertain parameter at the point of interest do not produce simple spatial shifts of the field, and the resulting field-shift-produced PDFs may have higher L_1 errors. Hence, these $D_{\min}$ calculations, which do not require knowledge of PDF$_c(A)$ and are already part of the approximate field-shift PDF technique, can indicate, at least approximately, when the final approximate PDF(A) result for that (r,z) location may be inaccurate.

Similar L_1 accuracy assessments were conducted for two uncertain parameters, channel depth H and mixed-layer sound speed c_1, using the idealized Fig. 1 sound channel. Figure 9 displays L_1 versus number of modes for the same frequencies, ranges, and parameters used for Fig. 7 with $\langle H\rangle$=50 m, $\sigma_H/\langle H\rangle$=0.005, $\langle c_1\rangle$=1500 m/s, $\sigma_{c_1}/\langle c_1\rangle$=0.001, and a receiver depth of 25 m. Here, 95% of the cases investigated yielded L_1 below 0.57. Again, high-L_1 outliers can exist where changes in one or both uncertain pa-

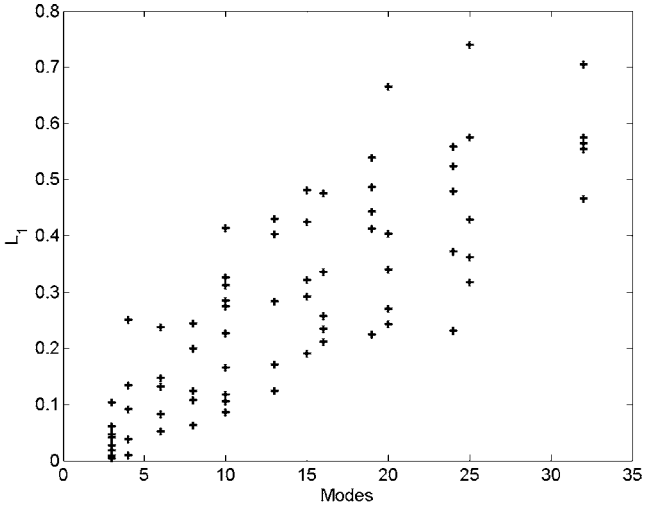


FIG. 9. L_1 vs the number of propagating modes in the Fig. 1 sound channels where the channel depth, H, and surface sound speed, c_1, are uncertain with $\langle H\rangle=50$ m, $\sigma_H/\langle H\rangle=0.005$, $\langle c_1\rangle=1500$ m/s, and $\sigma_{c_1}/\langle c_1\rangle=0.001$. All other parameters are as specified in Fig. 3.

rameters do not directly correspond to a spatial shift in the field, but these can be anticipated from their associated $D_{\min}$ values, as was illustrated for one uncertain parameter in Fig. 8.

A sample of three-uncertain-parameter L_1-results is shown in Fig. 10 for channels having the same uncertainties as in Fig. 9, but including the mixed-layer depth, d_1, as an additional uncertain parameter with $\langle d_1\rangle=10$ m and $\sigma_{d_1}/\langle d_1\rangle=0.01$. The frequencies, ranges, and receiver depth are identical to Fig. 7, and the bottom types are silt and sand with the same properties.

The results shown in Figs. 7, 9, and 10 can be condensed to show the overall accuracy of the field-shifting technique compared to the simple-linear approximation by plotting L_1 from the field-shifting technique versus the L_1 from simple-linear fitting. This plot is provided with logarithmic axes as Fig. 11 where results from one, two, and three uncertain parameters are plotted as pluses, circles, and triangles, respectively. The diagonal line has a slope of unity and indicates where field shifting and simple-linear fitting are equally accurate (i.e., they produce equal L_1 values). Results that fall below and to the right of this diagonal line indicate that the field-shifting technique is superior to simple linear fitting. Of the 600 cases plotted in Fig. 11, only 3% are better handled by the simple linear approximation, and for these cases the average difference in L_1 is 0.1. These results indicate that the spatial-shift-based approximate-PDF technique performs well in an idealized sound channel when compared to the simple linear approximation.

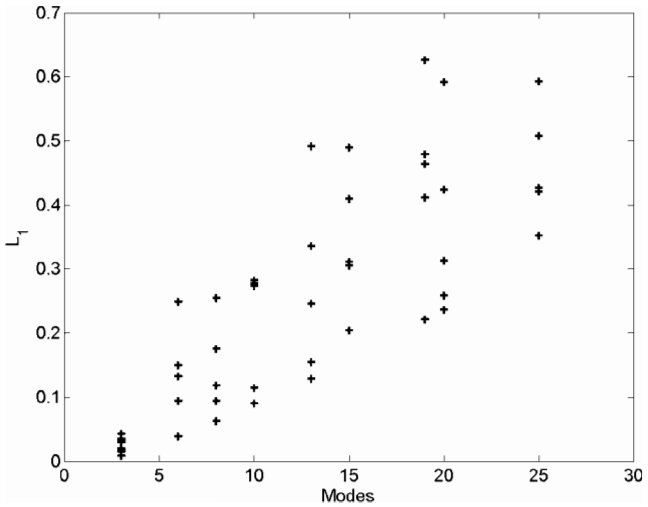


FIG. 10. Same as Fig. 9 except the mixed-layer depth, d_1, is also uncertain with $\langle d_1\rangle=10$ m and $\sigma_{d_1}/\langle d_1\rangle=0.01$. All other parameters are as specified in Fig. 3.

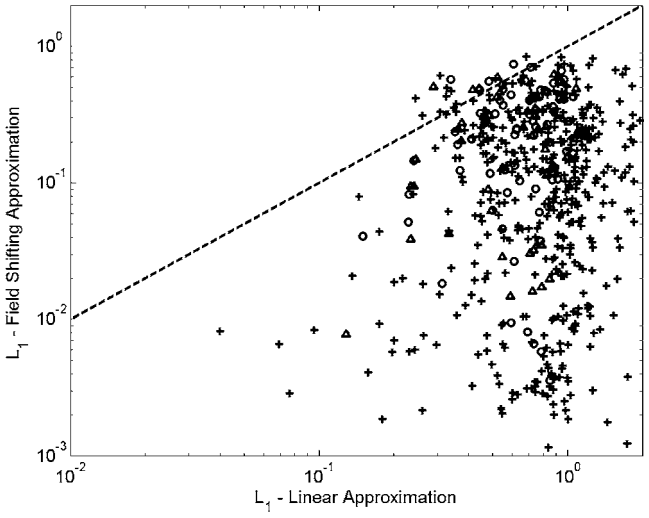


FIG. 11. Log-log plot of L_1 from the field-shifting technique vs L_1 from simple linear fitting. All previous results from one, two, and three uncertain parameters are plotted as pluses, circles, and triangles, respectively. The diagonal line has a slope of one and falls where the field-shifting and linear fitting PDF results are equal. Points below and to the right of this line indicate that the field-shifting technique is superior to simple linear fitting.

The final quantitative comparison provided in Fig. 12 is merely a demonstration that the field-shifting technique can

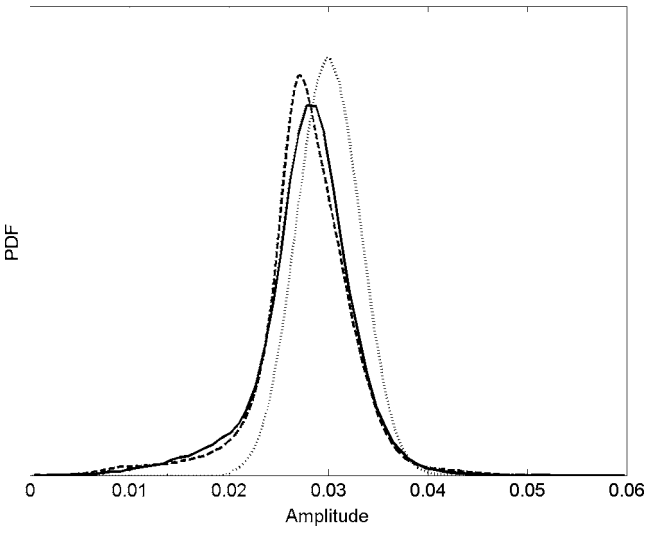


FIG. 12. Comparison of acoustic amplitude PDFs from field-shifting (dashed line), simple linear fitting (dotted line), and numerically converged direct sampling (solid line) for $f=400$ Hz at $(r,z)=(6.0\text{ km},25\text{ m})$ when all eight parameters of the Fig. 1 sound channel are uncertain. All other parameters and uncertainties are as specified in Fig. 3, except $\sigma_H/\langle H\rangle=0.0025$ and $\sigma_{c_1}/\langle c_1\rangle=0.0001$.

be successfully applied when all eight parameters specifying the Fig. 1 sound channel are uncertain. Here, the sound channel parameters have the same expected values and uncertainties as in Fig. 3, except $\sigma_H/\langle H\rangle=0.0025$ and $\sigma_{c_1}/\langle c_1\rangle=0.0001$. These two uncertainties were lowered so that all eight parameters contribute significantly to the final amplitude PDF. The L_1 values for the two approximate distributions shown in Fig. 12 are 0.12 for the field-shifting technique (dashed curve) and 0.45 for simple linear fitting (dotted curve). The field-shifting technique yields errors of 0.4% and 2% for the amplitude mean and standard deviation, respectively, while the linear approximation yields errors of 8% and 19%. Both approximate PDFs are based on nine field calculations while the numerically converged distribution (solid curve) to which they are compared is based on 1×10^6.

IV. SUMMARY AND CONCLUSIONS

This paper describes an approximate technique for determining the PDF of acoustic amplitude (A) when the acoustic amplitude is calculated from an acoustic field model having one or more uncertain environmental parameters as inputs. For N uncertain parameters, the inputs to the technique are the uncertain-parameter PDFs, and $N+1$ field calculations: one reference calculation without parametric variation and N sensitivity-assessment calculations at close-by values of each uncertain parameter. The technique is based on finding locally optimum spatial shifts between the calculated reference field and each sensitivity-assessment field so that the effects of arbitrary variations in uncertain input parameters can be approximated by easily calculated spatial shifts in the $N+1$ field calculations. The technique can be used with any suitable acoustic propagation model. It addresses the problem of quantifying acoustic field uncertainty in computational underwater acoustics and offers a less computationally demanding approach than direct numerical sampling or more sophisticated polynomial chaos techniques. This optimum-shift-based approximate-PDF technique was numerically tested in a variety of idealized range-independent shallow-water sound channels, and it appears to be potentially useful based on its relatively high accuracy from relatively few field calculations.

The following (enumerated) conclusions are drawn from this research effort. (1) Given that real time uncertainty analysis may be a valuable asset for sonar-based tactical decision aids, the approximate-PDF technique described here may be fast and accurate enough to be implemented in real time sonar applications. Furthermore, its simplicity may allow it or the optimum-shift concept to be utilized to save computational effort in environmental inversion, especially when nonlinear parametric sensitivity is needed for multiple parameters. (2) When interpreted properly, the L_1 error-norm is a useful overall indication of approximate PDF accuracy. Here, L_1 values up to 0.5 indicate converged-to-approximate PDF matching within engineering accuracy; L_1 of 0.5 typically implies up to 2 dB error for the standard deviation. (3) The primary limitation of the technique is that at some locations within a sound channel the field variations produced by changes in environmental parameters are not well described by a spatial shift, as stated in Dosso *et al.* (2007b). However, calculations completed as part of the field-shifting technique can reveal when high L_1 is likely. Extension of the field-shifting technique to acoustic phase and more geometrically complex environments awaits future research effort.

Furthermore, as a final comment, it must be noted that the parameter uncertainties for which the field-shifting technique has been developed and tested are relatively small in absolute terms, yet such uncertainties are large enough to produce nonlinearity in the acoustic amplitude versus uncertain-parameter curves. Thus, the lasting contribution from this work may be Eq. (7) or (11), the formulas for estimating nonlinear sensitivity to 1 or N parameters from only 2 or $N+1$ field calculations, respectively. The actual PDF production part of the technique may be less valuable because any acoustic uncertainty calculation that starts from typical ocean environmental parameter uncertainties may return a PDF(A) that is too broad to justify further use of a high-fidelity propagation code. In this circumstance where environmental uncertainties are high, environmental assessment and inversion to reduce the input parameter uncertainties is likely a requisite first step.

ACKNOWLEDGMENTS

This research effort was sponsored by the Office of Naval Research, Award Reference No. N00014-05-1-0243.

Brekhovskikh, L. M., and Lysanov, Yu. P. (**1991**). *Fundamentals of Ocean Acoustics*, 2nd ed. (Springer-Verlag, Berlin).
Colosi, J. A., and ATOC Group (**1999**). "A review of recent results on ocean acoustic wave propagation in random media: Basin scales," IEEE J. Ocean. Eng. **24**, 138–155.
Creamer, D. B. (**2006**). "On using polynomial chaos for modeling uncertainty in acoustic propagation," J. Acoust. Soc. Am. **119**, 1979–1994.
Dosso, S. E. (**2003**). "Environmental uncertainty in ocean acoustic source localization," Inverse Probl. **19**, 419–431.
Dosso, S. E., Giles, P. M., Brooke, G. H., McCammon, D. F., Pecknold, S., and Hines, P. C. (**2007a**). "Linear and nonlinear measures of ocean acoustic environmental sensitivity," J. Acoust. Soc. Am. **121**, 42–45.
Dosso, S. E., Morley, M. G., Giles, P. M., Brooke, G. H., McCammon, D. F., Pecknold, S., and Hines, P. C. (**2007b**). "Spatial field shifts in ocean acoustic environmental sensitivity analysis," J. Acoust. Soc. Am. **122**, 2560–2570.
D'Spain, G. L., and Kuperman, W. A. (**1999**). "Application of the waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468.
D'Spain, G. L., Murray, J. J., Hodgkiss, W. S., Booth, N. O., and Schey, P. W. (**1999**). "Mirages in shallow water matched field processing," J. Acoust. Soc. Am. **105**, 3245–3265.
Finette, S. (**2005**). "Embedding uncertainty into ocean acoustic propagation models," J. Acoust. Soc. Am. **117**, 997–1000.
Finette, S. (**2006**). "A stochastic representation of environmental uncertainty and its coupling to acoustic wave propagation in ocean waveguides," J. Acoust. Soc. Am. **120**, 2567–2579.
Gerstoft, P., Huang, C.-F., and Hodgkiss, W. S. (**2006**). "Estimation of transmisson loss in the presence of geoacoustic inversion uncertainty," IEEE J. Ocean. Eng. **31**, 299–307.
Gerstoft, P., and Mecklenbräuker, C. F. (**1998**). "Ocean acoustic inversion with estimation of a posteriori probability distributions," J. Acoust. Soc. Am. **104**, 808–819.
Haralabus, G., Premus, V., Alexandrou, D., Nolte, L. W., and Richardson, A. M. (**1993**). "Source localization in an uncertain acoustic scattering environment," J. Acoust. Soc. Am. **94**, 3379–3386.
Heaney, K. D., and Cox, H. (**2006**). "A tactical approach to environmental

uncertainty and sensitivity," IEEE J. Ocean. Eng. **31**, 356–367.
James, K. R., and Dowling, D. R. (**2005**). "A probability density function method for acoustic field uncertainty analysis," J. Acoust. Soc. Am. **118**, 2802–2810.
Jensen, F., Kuperman, W., Porter, M., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (AIP, New York).
Kim, S., Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Edelmann, G. F., and Akal, T. (**2003**). "Robust time reversal focusing in the ocean," J. Acoust. Soc. Am. **114**, 145–157.
LePage, K. D., and McDonald, B. E. (**2006**). "Environmental effects of waveguide uncertainty on coherent aspects of propagation, scattering, and reverberation," IEEE J. Ocean. Eng. **31**, 413–420.
Lin, Y.-T., Chen, C.-F., and Lynch, J. F. (**2006**). "An equivalent transform method for evaluating the effect of water-column mismatch on geoacoustic inversion," IEEE J. Ocean. Eng. **31**, 284–298.
Livingston, E. S., Goff, J. A., Finette, S., Abbot, P., Lynch, J. F., and Hodgkiss, W. S. (**2006**). "Guest editorial capturing uncertainty in the tactical ocean environment," IEEE J. Ocean. Eng. **31**, 245–248.
Papoulis, A. (**1965**). *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York), Chap. 5.
Porter, M., and Reiss, E. L. (**1984**). "A numerical method for ocean-acoustic normal modes," J. Acoust. Soc. Am. **76**, 244–252.
Sha, L., and Nolte, L. W. (**2005**). "Effects of environmental uncertainties on sonar detection performance prediction," J. Acoust. Soc. Am. **117**, 1942–1953.
Shang, E. C., and Wang, Y. Y. (**1991**). "Environmental mismatch effects on source localization processing in mode space," J. Acoust. Soc. Am. **89**, 2285–2290.
Shorey, J. A., Nolte, L. W., and Krolic, J. L. (**1994**). "Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments," J. Comput. Acoust. **2**, 285–314.
Sibul, L. H., Coviello, C. M., and Roan, M. J. (**2004**). "Detection of high frequency sources in random/uncertain media," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderious, and W. A. Kuperman (AIP, Melville, NY), pp. 237–244.
Stone, L., and Osborne, B. (**2004**). "Effect of environmental prediction uncertainty on target detection and tracking," Proc. SPIE **5428**, 58–69.
Wang, H., Yip, L., Yao, K., and Estrin, D. (**2004**). "Lower bounds of localization uncertainty in sensor networks," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 3, pp. iii–917–20.

Determination of thickness and elastic constants of aluminum plates from full-field wavelength measurements of single-mode narrowband Lamb waves

J. Luis Deán, Cristina Trillo, Ángel F. Doval, and José L. Fernández
Departamento de Física Aplicada, Universidade de Vigo, E. T. S. Enxeñeiros Industriais, Campus Universitario, E36310 Vigo, Spain

(Received 12 December 2007; revised 30 April 2008; accepted 26 May 2008)

A method based on fitting the theoretical dispersion curves of Lamb waves to experimental data is presented to determine the thickness and two independent elastic constants of aluminum plates a few millimeters thick. The waves are generated by means of the wedge method using a narrowband source, selecting the wedge angle and the acoustic frequency f so that mainly one mode is excited. A self-developed pulsed electronic speckle pattern interferometry system renders a two dimensional map of the out-of-plane acoustic displacement field at the plate surface, which allows an accurate measurement of the acoustic wavelength λ_1. For any mode, the relation between λ_1 and f depends on the three unknown parameters, so at least three experimental measurements (λ_{1i}, f_i) with different frequencies and/or different modes are required to calculate them. The suitability of different Lamb modes to determine each parameter when the others are known is studied, as well as the conditions that the experimental set of values must fulfill to calculate all three parameters. Numerous Lamb modes at different frequencies are generated in each plate, and a fitting is made based on the minimization of the error function, resulting in an accuracy better than 1%.
 [DOI: 10.1121/1.2945707]

PACS number(s): 43.35.Cg, 43.35.Pt, 43.35.Sx, 43.35.Zc [TDM] Pages: 1477–1489

I. INTRODUCTION

Ultrasonics is arguably one of the most interesting means for the determination of the adiabatic elastic constants of materials. Its accuracy, nondestructive nature, and well-proved validity are key advantages that make the use of ultrasound adequate to the elastic characterization of a sample.[1] The procedure consists in the measurement of the phase and/or group velocities of specific acoustic waves, being the longitudinal and shear bulk wave velocities, c_L and c_T respectively, usually employed. These two velocities suffice to acoustically characterize an isotropic and homogeneous material, i.e., to determine the behavior of all the possible acoustic waves within any sample of such material. Furthermore, Poisson's ratio ν can be calculated, as well as other adiabatic elastic constants such as $H=E/\rho$, λ_L/ρ, or μ_L/ρ. Thereby, with the additional knowledge of the mass density ρ, Young's modulus E or the Lamé constants λ_L and μ_L can be obtained.[2]

Lamb waves are a kind of guided waves that propagate along plates with stress-free boundaries as a consequence of the superposition of longitudinal and shear waves reflected in the plate surfaces.[3] They have been widely used in the inspection of plates and similar elements, including shell structures, since for this type of specimens Lamb waves present important advantages compared to bulk waves. Among these advantages we can highlight their capability to propagate over large distances with low attenuation (thus allowing a wider inspection area than bulk waves), their availability in a broad range of frequencies (not limited by the wavelength/thickness quotient as it is the case in bulk waves), and their applicability to measure directional properties in anisotropic media. However, probably the most appealing feature of Lamb waves is the existence of many usable dispersive modes, so that the field distribution depends on mode and frequency, and a proper selection allows one to perform selective and/or sensitive measurements.

Pointwise detection systems are usually employed to collect acoustic signals using either contact or noncontact transducers[4] (piezoelectric transducers, electromagnetic acoustic transducers (EMATs) air-coupled transducers, pointwise laser interferometers, etc). These detectors provide a single-channel signal that integrates information about the temporal evolution of a surface zone, so that several measurements at different locations are required to measure the phase or group velocities of the acoustic wave. On the other hand, full-field interferometric techniques such as electronic speckle pattern interferometry (ESPI) cover a large area of the surface in each measurement, allowing a faster evaluation of the spatial features of the wave. Indeed, the displacement maps that ESPI yields allow to visualize the Lamb wave fronts and to detect perturbations due to flaws or other heterogeneities.[5,6]

In our domain of interest of macroscopic structural elements, several authors have employed Lamb waves to calculate the thickness and/or the elastic constants of isotropic plates with thicknesses ranging from 0.2 to 5 mm, and the use of both broadband[7–15] and narrowband signals has been reported.[16–18] Broadband Lamb waves contain information over a wide area of the frequency spectrum; nonetheless, a processing of this information is required in order to assess the influence of each mode at each frequency. The most usual processing techniques are referred to in Ref. 18. In

general, the processing procedure yields the phase or group velocities of the excited modes versus frequency, i.e., the dispersion curves, within the frequency range. Then, the desired elastic constants and/or the plate thickness are obtained as the parameters of the theoretical model that best fit the experimental data. Narrowband signals, although containing less information, provide a higher signal-to-noise ratio and, even when they are multimode and need to be processed, yield more accurate measurements than those of the broadband type. The frequency range depends on the plate thickness, because Lamb modes are described in terms of the frequency-thickness product. In thin plates, high frequency ultrasound is needed, usually generated with short laser pulses.[7,19,20]

Since, in isotropic and homogeneous materials, the Lamb wave phase and group velocities depend on the plate thickness and on two independent constants (that can be the bulk wave velocities, Poisson's ratio, or other elastic constants divided by the mass density), the experimental data can be used to calculate one of these parameters,[7,8,12,13,18,20] two of them (generally the elastic constants),[9,11,14–17] or all three.[10,19] In other cases, the dispersion curves of Lamb waves depend on more constants and more parameters can be fitted; that is the case of anisotropic materials,[21–23] bilayered films,[24,25] or coating layers on a substrate.[26,27]

To calculate the plate constants, least-squares fitting criteria are generally applied, i.e., a procedure is followed in order to minimize the summation of the squared differences between the experimental data and the theoretical curves. Numerical procedures, usually based on the simplex algorithm,[8,11,13,22,23,26,28] are employed because of the complicated nature of the equations. Furthermore, it can be useful to make a graphical representation of the difference between experimental and theoretical data as a function of the constants, but this is only possible when intending to calculate two parameters at the most.[9,16,18] When dealing only with the $A0$ and/or $S0$ Lamb modes, some authors adopt a different approach based on substituting approximations for the exact expression for the Rayleigh–Lamb dispersion curves,[7,10,19,20,24] which greatly simplifies the fitting procedure.

In this paper we measure independently the frequency f and the wavelength λ_1 of single-mode Lamb waves in aluminum plates, so that the phase velocity can be obtained as their product.[29] Then, we calculate the material bulk wave velocities and the plate thickness by fitting the elastic model to these experimental data. To measure the wavelength, we employ a full-field optical technique that belongs to the group of coherent techniques under the common denomination of ESPI or TV holography. The organization of the paper is as follows. In Sec. II a brief review of the well-known theory of Lamb waves is made in order to introduce notation. Furthermore, we present an analysis of the sensitivity to the determination of each parameter as a function of mode and frequency and also study in which cases the measurement of only three experimental points (λ_{1i},f_i) suffices to characterize the plate material and thickness. Such study is useful to select the best modes and frequencies to achieve our goal; however, we must also take into account the feasibility of the excitation of single modes, because, in the case that several modes coexist, a beating is produced that raises errors in the measurement of the individual wavelength of each mode.[18] In Sec. III we make a comprehensive description of the experimental procedure, placing particular emphasis on the measurement of the mode wavelength by means of ESPI and in its experimental uncertainty. Finally, in Sec. IV, we explain the fitting procedure of the parameters and display the results.

The method herein presented permits the characterization of an isotropic plate, i.e., the determination of the thickness and the bulk wave velocities. All three parameters are obtained from the experimental data, and the analysis of their uncertainties permits to conclude that the accuracy of the method is rather high. Furthermore, the detection is performed remotely (the optical head is several meters apart from the plate), and although the wave generation is currently achieved by means of a contact method, we expect to implement a laser generation system in the future, and so to develop a completely remote and sensitive technique for the determination of the thickness and the elastic constants of plate materials.

II. THEORY

A. Lamb waves

The well-known Rayleigh–Lamb complex equation establishes the dispersion relation of any possible symmetric (S) or antisymmetric (A) Lamb mode. It is deduced assuming a time-harmonic dependence, resulting in adimensional form[2]

$$F_A^S = \frac{\tan(\pi/2)\sqrt{\gamma^2\chi^2-\xi^2}}{\tan(\pi/2)\sqrt{\gamma^2-\xi^2}} + \left[\frac{4\xi^2\sqrt{\gamma^2\chi^2-\xi^2}\sqrt{\gamma^2-\xi^2}}{(2\xi^2-\gamma^2\chi^2)^2}\right]^{\pm 1} = 0, \tag{1}$$

where

$$\xi = \frac{2hk_1}{\pi} \quad \text{is the normalized wave number,} \tag{2}$$

$$\gamma = \frac{2h\omega}{\pi c_L} \quad \text{is the normalized frequency,} \tag{3}$$

$$\chi = \frac{c_L}{c_T} \quad \text{is the velocity ratio,} \tag{4}$$

and k_1 is the Lamb wave number (for a wave propagating along the x_1 direction), $2h$ is the plate thickness, and $\omega = 2\pi f$ is the circular frequency.

For each frequency, either real or complex values of the wave number that satisfy the Rayleigh–Lamb equation can be found. Propagating Lamb waves correspond to real values of the wave number, in a way that a dispersion relation can be written for each mode as

$$k_1 = k_1(f, c_L, c_T, 2h). \tag{5}$$

The real solutions of Eq. (1) for the first five symmetric and antisymmetric Lamb modes are plotted in Fig. 1, calculated for an aluminum plate having $c_L = 6330$ m/s and

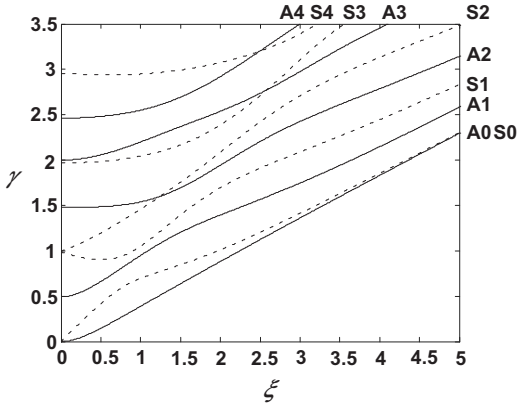


FIG. 1. Representation of the frequency spectrum of antisymmetric (continuous lines) and symmetric (dashed lines) Lamb modes in a plate having c_L=6330 m/s and c_T=3117 m/s.

c_T=3117 m/s. It shows the frequency spectrum, where the normalized frequency is represented against the normalized wave number.

The instantaneous out-of-plane (in the x_3 direction) acoustic displacement $u_{l,3}$ at a point $\mathbf{x}=(x_1,x_2)$ of the plate surface due to a particular plane narrowband Lamb mode l is given by

$$u_{l,3}(\mathbf{x},t) = \mathrm{Re}[\hat{u}_{l,3}(\mathbf{x},t)] = \mathrm{Re}[u_{l,3m}(x_1)\exp\{j(k_{l,1}x_1 - \omega_l t + \varphi_{l,30})\}], \quad (6)$$

where the relation between the wave number $k_{l,1}$ and the circular frequency ω_l of the mode l is detailed in the curves of Fig. 1. Henceforth, for simplicity we will omit the subscript l, unless it is necessary to identify a particular mode.

B. Best frequencies and modes

In Sec. II A we described how the dispersion relation, Eq. (5), depends on the plate thickness and on the velocities of the longitudinal and shear waves.

Therefore, the theoretical model has three degrees of freedom and, consequently, at least three experimental points (k_{1i},f_i), being k_{1i} and f_i the experimental values of the wave number and frequency, are needed to acoustically characterize the plate. Instead of the directly measured wavelength λ_1, we have selected the wave number k_1 as the experimental quantity for fitting the model parameters since, as we explain in Sec. III E, the uncertainty in k_1 is approximately constant for all the frequencies and modes used in the experiments, and so the fitting procedure becomes easier.

With just one experimental measurement, any parameter c_L, c_T, or $2h$ can be obtained if the rest of them are known, as the value of such parameter must verify the following equation:

$$k_1(f_i,c_L,c_T,2h) - k_{1i} = 0. \quad (7)$$

The accuracy in the determination of any of the parameters depends on how its variation affects the variation of k_1, so we define the sensitivity estimators[16] s_L, s_T, and s_h as

$$s_L(f,c_L,c_T,2h) = \frac{\partial k_1(f,c_L,c_T,2h)}{\partial c_L}, \quad (8)$$

$$s_T(f,c_L,c_T,2h) = \frac{\partial k_1(f,c_L,c_T,2h)}{\partial c_T}, \quad (9)$$

$$s_h(f,c_L,c_T,2h) = \frac{\partial k_1(f,c_L,c_T,2h)}{\partial (2h)}. \quad (10)$$

Then, higher values of a sensitivity estimator correspond to smaller values of the uncertainty in the corresponding parameter for a given uncertainty in k_1. The variation of these estimators versus the frequency-thickness product for the first five symmetric and antisymmetric modes is displayed in Fig. 2.

When the aim is to determine more than one parameter, high values of the sensitivity estimators do not suffice to render a good accuracy in the measurements. For example, in principle, two measurements, (k_{1i},f_i) with $i=1,2$, should suffice to calculate c_L and c_T in a plate of known thickness. Each individual measurement cannot render a solution for both velocities, but a relation between them corresponding to the set of values that verifies Eq. (7), which can be regarded as a curve $c_T=g_{Ti}(c_L)$, so that the intersection point (c_{L12},c_{T12}) of two curves represents the solution for c_L and c_T. Then, the accuracy in the result is higher when the slopes of the curves are rather different [Fig. 3(a)] than when they are similar [Fig. 3(b)] since errors in the experimental data (k_{1i},f_i) affect much more the localization of the intersection of the curves in the second case.

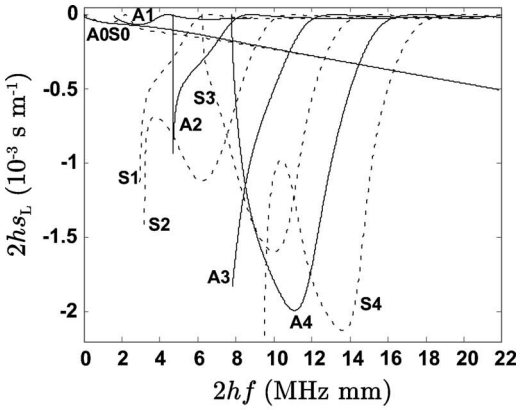
For calculating the slope of each curve given the thickness and the experimental data (k_{1i},f_i), we consider that Eq. (7) is verified along the curve, so that the differentiation of this expression yields

$$\frac{\partial k_1(f_i,c_L,c_T,2h)}{\partial c_L}dc_L + \frac{\partial k_1(f_i,c_L,c_T,2h)}{\partial c_T}dc_T = 0, \quad (11)$$

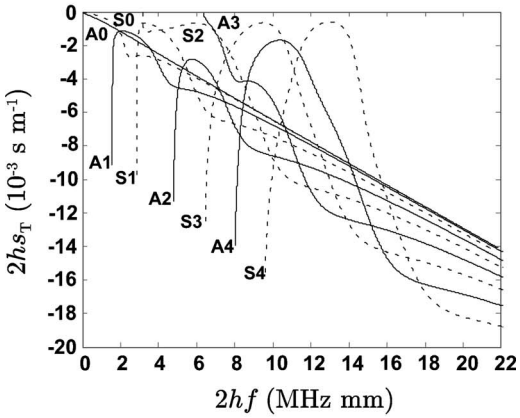
Therefore,

$$\frac{dc_T}{dc_L} = \frac{dg_{Ti}}{dc_L} = -\frac{\partial k_1(f_i,c_L,c_T,2h)/\partial c_L}{\partial k_1(f_i,c_L,c_T,2h)/\partial c_T} = -\frac{s_L(f_i,c_L,c_T,2h)}{s_T(f_i,c_L,c_T,2h)}. \quad (12)$$

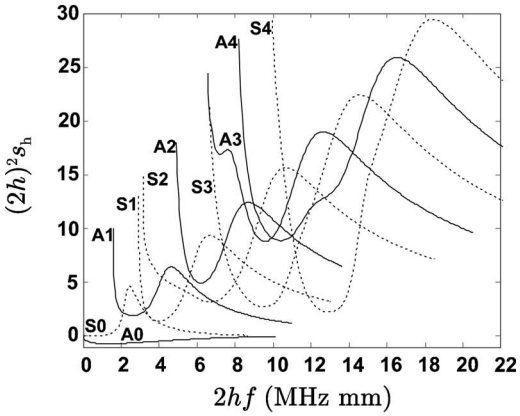
Equally well, the solution rendered by each individual measurement (k_{1i},f_i) can be represented as a surface $2h=g_{hi}(c_L,c_T)$ in a Cartesian system of three coordinates $(c_L,c_T,2h)$ when none of these three parameters is known, and at least three points $(i=1,2,3)$ are needed to obtain a solution $(c_{L123},c_{T123},2h_{123})$ as the intersection of such surfaces. Ideally, the most accurate solution is obtained if the three surfaces are normal to each other, so, to assess the capability of a given set of experimental points to yield an accurate solution, a criterion must be established in order to value how close we are to that optimal condition. For example, we can consider if the mixed product between the normals to the three surfaces has a high value. The normal direction to the surface i is given by the unit vector $\mathbf{n}_i/|\mathbf{n}_i|$, being $\mathbf{n}_i$ given by Eq. (13) as a function of the slopes of the curves $2h=g_{hi}(c_L,c_{T123})$ and $2h=g_{hi}(c_{L123},c_T)$ yielded by each experimental measurement (k_{1i},f_i) when the other parameter (c_{T123} or c_{L123}, respectively) is fixed. For a generic point of the surface i, we can write



(a)



(b)

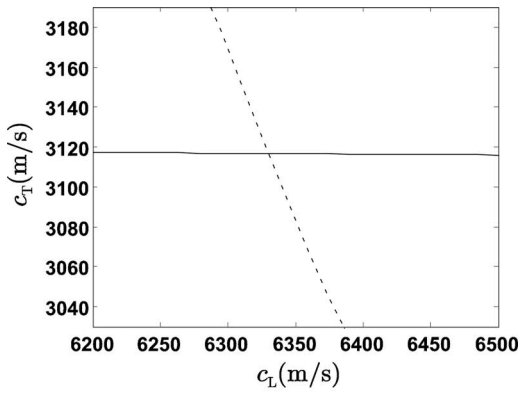


(c)

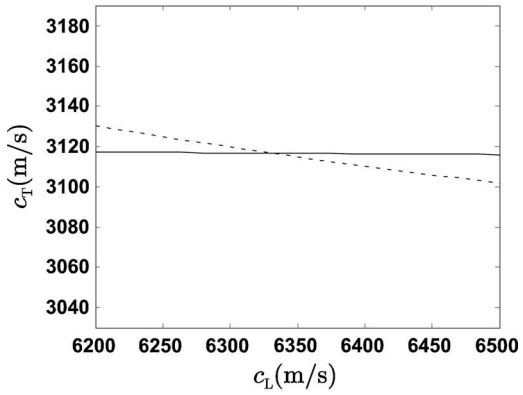
FIG. 2. Representation of the sensitivity estimators against the frequency-thickness product in a plate with c_L=6330 m/s and c_T=3117 m/s. The continuous lines correspond to the antisymmetric modes and the dashed lines to the symmetric ones. (a) Sensitivity estimator for c_L multiplied by $2h$. (b) Sensitivity estimator for c_T multiplied by $2h$. (c) Sensitivity estimator for $2h$ multiplied by $2h$ squared. The advantage of using the factors $2h$ and $2h$ squared is that the represented functions do not depend on $2h$.

$$\mathbf{n}_i = \left(-\frac{\partial g_{hi}(c_L, c_T)}{\partial c_L}, -\frac{\partial g_{hi}(c_L, c_T)}{\partial c_T}, 1 \right). \tag{13}$$

The normal vector $\mathbf{n}_i$ is given as a function of other slopes if the surfaces are expressed as $c_T = g_{Ti}(c_L, 2h)$ or $c_L = g_{Li}(c_T, 2h)$. Therefore, when the problem consists in finding two or three parameters, it is important to study the dependence of one or two, respectively, of the functions $\partial g_{hi}/\partial c_L$, $\partial g_{hi}/\partial c_T$, and $\partial g_{Ti}/\partial c_L$ on mode and frequency. Their values for the first five symmetric and antisymmetric Lamb modes are represented in Fig. 4.



(a)



(b)

FIG. 3. Each curve $c_T = g_{Ti}(c_L)$, with i=1,2,3, corresponds to the solution for the bulk velocities that satisfies Eq. (7) yielded by a particular experimental point (k_{1i}, f_i) when $2h$ is fixed, calculated for a plate with c_L =6330 m/s and c_T=3117 m/s. (a) Mode $S0$ at $2hf$=1.5 MHz mm (continuous line) and mode $S2$ at $2hf$=6 MHz mm (dashed line). (b) Same mode $S0$ as in (a) (continuous line) and mode $A1$ at $2hf$=4 MHz mm (dashed line).

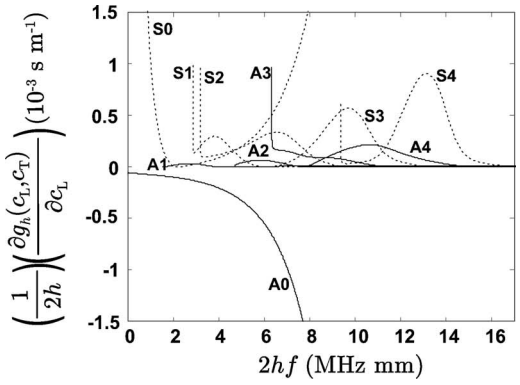
In any case, whenever possible, it is better to employ a large number of points not only because of the improvement in the accuracy but also to check the validity of the theoretical model. In our case, the fitting of the parameters was performed with all the single modes that we were able to measure in all the plates.

III. EXPERIMENTAL PROCEDURE

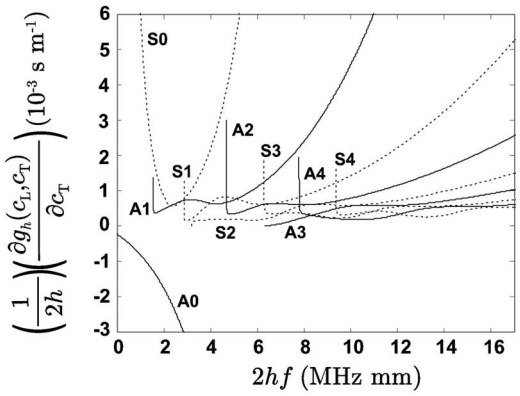
A. Generation of the Lamb waves

The test specimens are aluminum plates (alloy denomination EN AW-2017A-T4) with thicknesses between 1.5 and 5 mm. The plates are supported in a way close to the free-space condition, i.e., minimizing constraints at the surface; they simply rest on a horizontal board covered in velvet fabric.

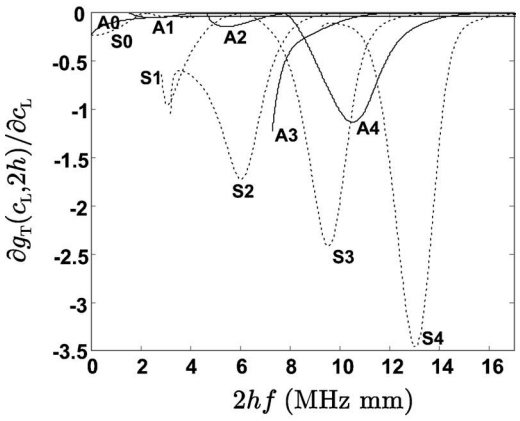
The generation of Lamb waves with out-of-plane amplitudes large enough to be detectable with our pulsed ESPI system can be performed with several methods: piezoelectric transducer, EMAT, laser, etc. Among them, laser generation seems to be the most interesting option because of the ad-



(a)



(b)



(c)

FIG. 4. Representation of the slopes of the curves yielded by a generic experimental point (k_{1i},f_i) when one parameter is fixed, against the frequency-thickness product, calculated for a plate with c_L=6330 m/s and c_T=3117 m/s. The continuous lines correspond to the antisymmetric modes and the dashed lines to the symmetric ones. (a) Slope of $2h$ with respect to c_L, divided by $2h$. (b) Slope of $2h$ with respect to c_T, divided by $2h$. (c) Slope of c_T with respect to c_L.

vantages derived from its capability of remote operation. However, the laser excitation of narrowband waves often requires a rather expensive equipment.[30–35]

In our case, we performed the excitation of the waves by means of the so-called prismatic coupling block method,[36] a contact technique in which an acoustic wave generated by a piezoelectric transducer (PZT) is coupled to the plate by means of a wedge (Fig. 5). The longitudinal wave that propagates through the wedge reaches the plate under a certain angle of incidence θ, so that it acts as an acoustical source

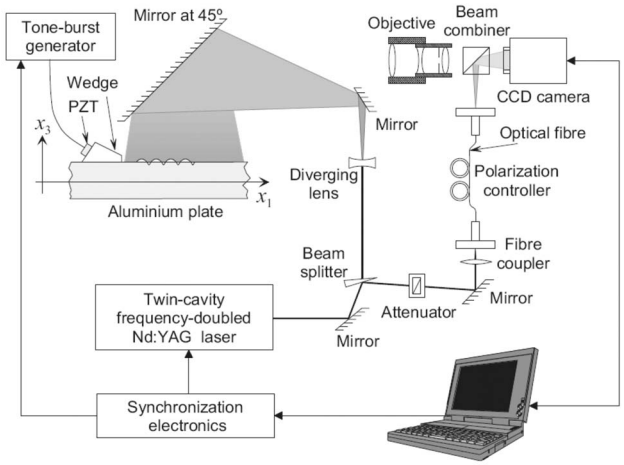


FIG. 5. Layout of the double-pulsed ESPI system.

with a given spatiotemporal profile in the contact area with the plate. For narrowband excitation, the spatial period of this source measured along the longitudinal coordinate x_1 on the plate surface (Fig. 5), namely, the wavelength λ_1 imposed by the wedge, is given by

$$\lambda_1 = \frac{\lambda_w}{\sin\theta}, \tag{14}$$

where λ_w is the wavelength of the longitudinal wave inside the wedge.

The efficiency in the generation of an isolated narrowband Lamb mode depends on many factors, but, once ensured that the imposed frequency and wavelength match the values of the desired mode, the efficiency is higher when the contact area wedge plate (that is, the excitation zone) is large compared to the Lamb wavelength.[36] In any case, other modes with wavelengths close to λ_1 are usually generated, even using monochromatic excitation.

In practice, apart from selecting good sensitivities and adequate slopes we must also consider the feasibility of exciting the mode isolated and not together with other modes that may introduce serious errors or even prevent the measurement of the wavelength. For example, Fig. 2 shows that the sensitivity estimators of some modes take high values when the experimental point is close to the cutoff frequency; however, a rather large excitation zone is required to cover several wavelengths and so to generate these modes efficiently, which is impractical in our plates.

Instead of using pure monochromatic excitation, we have utilized a narrowband source consisting in a long tone burst with constant amplitude and sinusoidal profile, having enough duration to ensure that Eq. (6) applies. Thereby, a unique value of the wavelength can be assigned to the Lamb wave train. On the other hand, we have limited the burst length so that the wave front has not reached any edge of the plate at the measurement instant; hence no acoustic absorbers at the edges are needed to avoid disturbing reflection phenomena (mode conversion, interference, etc.). Although these reflected waves could be removed from the measured displacements by processing the complete dynamic displace-

TABLE I. Calculated values of the thickness $2h$, the longitudinal wave velocity c_L, and the shear wave velocity c_T obtained from minimizing $\chi_{k_1}(c_L, c_T, 2h)$ with n experimental points. From them, the values of $H=E/\rho$, being E Young's modulus and ρ the mass density, and Poisson's ratio ν were calculated. The values of $2h$ and c_L obtained, respectively, with a micrometer and by means of the pulse-echo method are also given for reference purposes.

$2h$ (mm) (outside micrometer)	c_L (m/s) (pulse echo)	n	$2h$ (mm)	c_L (m/s)	c_T (m/s)	H (10^7 m^2/s^2)	ν
1.51 ± 0.01	⋯	34	1.511 ± 0.007	6354 ± 39	3113 ± 5	2.601 ± 0.009	0.3420 ± 0.0027
1.97 ± 0.01	⋯	38	1.968 ± 0.009	6348 ± 36	3109 ± 7	2.594 ± 0.011	0.3420 ± 0.0025
2.98 ± 0.01	6330 ± 30	46	2.976 ± 0.024	6378 ± 30	3123 ± 12	2.614 ± 0.017	0.3422 ± 0.0025
4.03 ± 0.01	6330 ± 30	67	4.013 ± 0.032	6371 ± 30	3107 ± 10	2.597 ± 0.015	0.3441 ± 0.0024
5.00 ± 0.01	6315 ± 30	70	4.996 ± 0.033	6382 ± 25	3112 ± 8	2.604 ± 0.012	0.3440 ± 0.0020

ment field, our technique can be performed with only one displacement map, exploiting the possibility to record clean, transient traveling waves.

B. Measurement of the acoustic frequency

The burst is, in principle, long enough to assume that a unique frequency f can be assigned to the wave train, corresponding to the central frequency of the signal generated by the piezoelectric driving electronics. Nevertheless, we experimentally checked the width of the frequency spectrum by measuring the out-of-plane displacement of a surface point with a Michelson speckle interferometer,[29] resulting that the waves are narrowband enough to satisfy our assumption.

C. Measurement of the thickness and the longitudinal bulk wave velocity

In order to validate the values obtained by fitting the theoretical model to the experimental data, we have directly measured $2h$ and c_L by means of an outside micrometer and the classical pulse-echo method, respectively (Table I). Furthermore, as we can see in Fig. 2, s_L is, in general, much lower than s_T, so, provided the plate thickness is known, it could also be possible to obtain a good accuracy in c_L and c_T by combining the measurement of a Lamb mode wavelength and the direct measurement of c_L, by using a fitting procedure similar to the one described in Ref. 18. The width of a single pulse in the frequency range of the PZTs at our disposal (maximum operating frequency of about 3 MHz) is in the order of the time of flight along the thickness direction, so that the successive echoes are partially overlapped. Such overlapping hampers the measurement of the longitudinal wave velocity, and although it could be possible to obtain c_L in all the plates by employing more sophisticated methods,[37–39] we only performed the measurement in the thick ones, as we consider it enough for validating purposes.

D. Measurement of the wavelength

The wavelength of the Lamb modes is obtained from the out-of-plane displacement maps rendered by our double-pulsed ESPI system, whose layout is depicted in Fig. 5.

ESPI is a full-field interferometric technique based on the electronic recording of primary correlograms, which are formed as a consequence of the interference between a reference beam and an object beam scattered back by the part to be measured.[40] In our case, the variable part of the object optical phase difference $\phi_o(\mathbf{x},t)$, i.e., the component of the optical phase depending on the change in the position of the object points, is related to the instantaneous out-of-plane displacement field $u_3(\mathbf{x},t)$ at the plate surface as follows:

$$\phi_o(\mathbf{x},t) = -\frac{4\pi}{\lambda} u_3(\mathbf{x},t), \tag{15}$$

where λ is the wavelength of the laser.

We use a pulsed laser to acquire two correlograms with a short time delay between them. Then, we apply a processing procedure based on the spatial Fourier transform method[41] (SFTM) to obtain the optical phase change $\Delta\Phi(\mathbf{x},t)$ between the recording instants, which is related to the object optical phase difference $\phi_o(\mathbf{x},t)$ and hence to the out-of-plane displacements. The highest sensitivity under narrowband excitation is achieved when the delay between the two recording instants is set to an odd number of half periods of the acoustic wave. In that case, the instantaneous optical phase-change map for a single-mode Lamb wave can be approximately expressed as follows:

$$\Delta\Phi(\mathbf{x},t_1) = \frac{8\pi}{\lambda} u_3(\mathbf{x},t_1), \tag{16}$$

where t_1 is the initial instant of the measurement (first laser pulse), which can be conveniently set by means of the synchronizing electronics.

Equation (16) is not valid at the front and trailing edges of the wave train,[42] since the wave train shifts between the two recording instants due to its traveling wave character. Furthermore, Eq. (6) is not valid in this region either. Nevertheless, we used long tone bursts of about 90 cycles and the delay between the two recording instants is typically three or five half periods. Therefore, the region in which Eq. (16) is not valid is small compared to the length of the wave train, and whenever possible, we leave this region outside the field of view.

Typical examples of the resulting optical phase-change maps are shown in Fig. 6(i), corresponding to waves with amplitudes in the order of 5 nm, which are clearly distinguished with our system.[41] These maps serve to measure the wavelength of a monomode Lamb wave with a rather good precision. However, we applied a second evaluation stage described in Ref. 41 also based on the SFTM, which

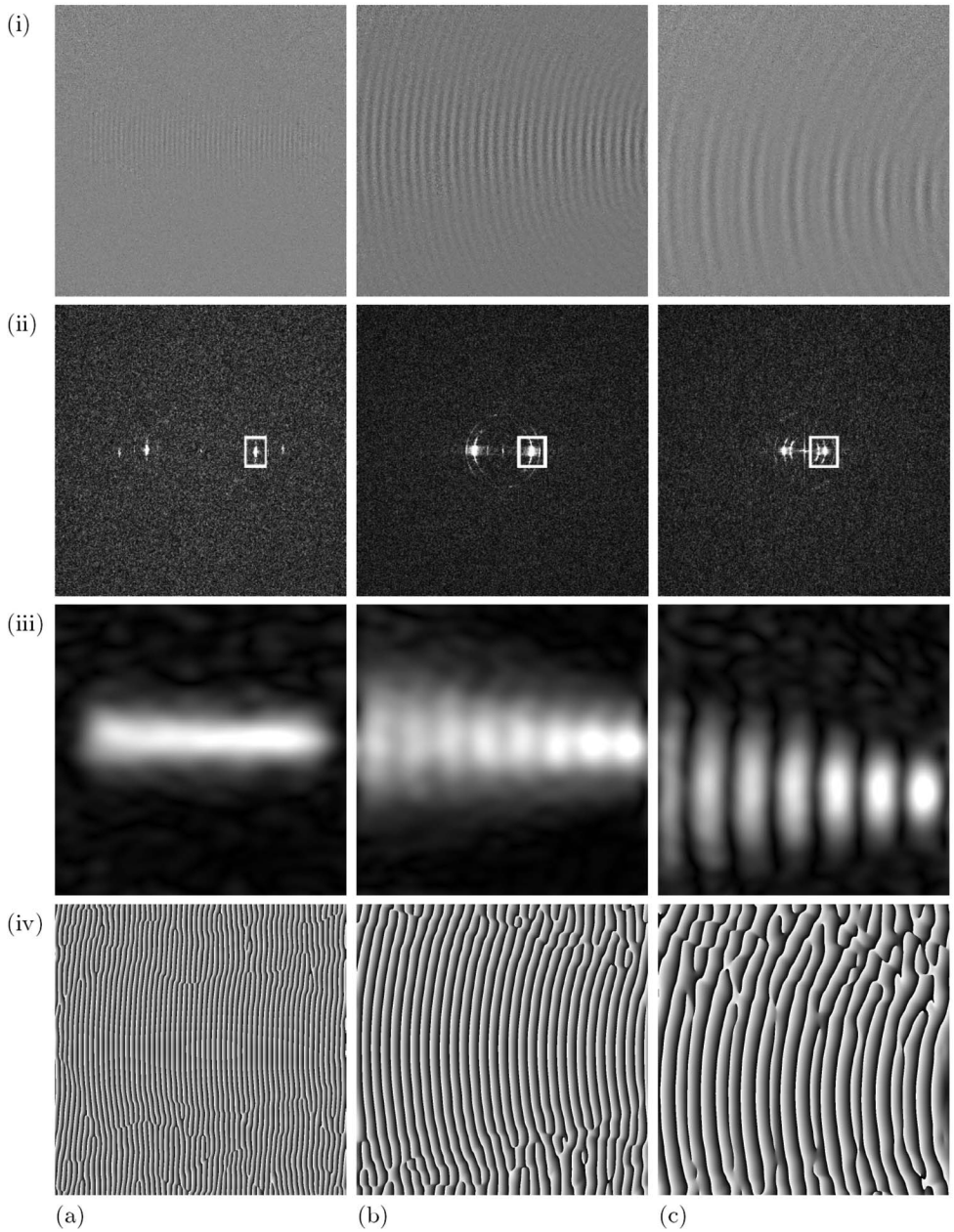


FIG. 6. Out-of-plane displacement maps of several Lamb waves at a given instant t_1 measured with our double-pulsed ESPI system. (i) Instantaneous optical phase-change map, $\Delta\Phi(\mathbf{x},t_1)$. (ii) Fast Fourier transform of the optical phase-change map indicating the position of the filter. (iii) Acoustic amplitude, $u_{3m}(\mathbf{x})$. (iv) Acoustic phase, $\arg[\hat{u}_3(\mathbf{x},t_1)]$. (a) Mode $A2$ at 2.222 MHz in an aluminum plate 4.03 mm thick. The actual size of the field of view is 111×111 mm^2. (b) Modes $S0$ and $A0$ at 0.909 MHz in an aluminum plate 2.98 mm thick, being the amplitude of mode $S0$ much larger than the amplitude of mode $A0$. The actual size of the field of view is 116×116 mm^2. (c) Modes $S1$ and $S2$ with approximately the same amplitude at 1.053 MHz in an aluminum plate 5.00 mm thick. The actual size of the field of view is 111×111 mm^2.

basically comprises three steps. First, the fast Fourier transform (FFT) of the instantaneous optical phase-change map is calculated. Afterward, as shown in Fig. 6(ii), a bandpass filter is applied to isolate one of the side lobes that contains the spatial frequencies of interest. Finally, the inverse FFT is computed. The final result for a single-mode Lamb wave is a complex map $\Delta\hat{\Phi}(\mathbf{x},t_1)$ whose modulus [Fig. 6(iii)] is proportional to the acoustic amplitude $u_{3m}(\mathbf{x})$ and whose argument [Fig. 6(iv)] is equal to the total acoustic phase $\arg[\hat{u}_3(\mathbf{x},t_1)]$. Taking $t_1=0$, the final complex map for a single-mode narrowband Lamb wave is given by

$$\Delta\hat{\Phi}(\mathbf{x}) = \frac{4\pi}{\lambda}u_{3m}(\mathbf{x})\exp\{j\varphi_{3m}(\mathbf{x})\}, \tag{17}$$

being $\varphi_{3m}(\mathbf{x})=k_1x_1+\varphi_{30}$.

The application of this method has several advantages. For instance, the bandpass filtering reduces substantially the speckle noise still present in the optical phase-change maps. Furthermore, the resulting acoustic amplitude maps can be used to check the single-mode nature of the waves. As a matter of fact, when there is more than one mode propagating along the plate, a spatial beating is produced that is

clearly detected in these maps. On the other hand, in the case of a single-mode wave, the amplitude u_{3m} is constant throughout the plate or varies smoothly along the propagation direction. In Fig. 6(c)(iii) a strong beating between two modes is observed while in Fig. 6(a)(iii) only one mode is detected. Figure 6(b)(iii) shows an example in which a soft beating is perceived, i.e., there is a second mode interfering whose amplitude is much smaller than the amplitude of the main mode. In this case, as we will explain in the Appendix, the presence of the second mode does not prevent the wavelength of the main mode from being measured.

The acoustic phase maps of Fig. 6(iv) provide an accurate means to directly measure the wavelength λ_1 of a single mode. The acoustic phase when a single mode propagates along the x_1 direction is given by Eq. (6). If the image sensor is aligned with the pixel rows parallel to x_1, the acoustic phase increment between two points P_i and P_f of the same row of the image is as follows:

$$\Delta\varphi_{3m} = \frac{k_1 N_{if}}{M}, \tag{18}$$

where M is the optical magnification expressed as the number of pixels per unit of length at the plate surface, and N_{if} is the distance in pixels between the points.

To measure the wavelength, we typically employ marking points P_i and P_f where a transition from $-\pi$ to π occurs in the phase map, because they are easily identified. Then, the actual phase difference between the points is $\Delta\varphi_{3m} = 2\pi N$, N being the number of acoustic wave periods in between, resulting in

$$\lambda_1 = \frac{2\pi}{k_1} = \frac{N_{if}}{MN}. \tag{19}$$

As we said in Sec. III A, the wavelengths of the different modes that can be excited with a certain wedge at a given frequency are, in general, rather close to the wavelength imposed by the wedge. Therefore, their spatial frequencies are rather similar, and this fact prevents the isolation of one of the modes in the Fourier spectrum, i.e., it is not feasible to apply a bandpass filter in order to remove the nondesired modes without affecting the one to be measured. Nevertheless, in some cases two modes with relatively different spatial frequencies can be generated and then it is possible to isolate each one and measure both wavelengths in the final maps. Figure 7(i) shows the instantaneous optical phase-change map produced by the simultaneous presence of modes $A1$ and $S0$. In Figs. 7(iii) and 7(iv) the amplitude and phase are displayed when filtering each one of the modes [Figs. 7(a) and 7(b)] or when considering both together [Fig. 7(c)].

E. Uncertainties in the measurements

The frequency uncertainty of the PZT electronic driver was estimated to be 5 kHz from the manufacturer's specifications; furthermore, the central frequency values in the spectra of the signals measured with the Michelson interferometer are consistent with this estimation.[29] Equally, the uncertainties in the determination of the thickness and the longitudinal wave velocity are approximately 0.01 mm and 30 m/s, respectively.[18]

However, the main concern is related to the determination of the uncertainty in the Lamb mode wave number k_1. The value of k_1 for a particular mode is obtained from Eq. (18); hence, its uncertainty $u(k_1)$ depends on the uncertainties in the distance N_{if}, the magnification M, and the acoustic phase increment $\Delta\varphi_{3m}$. As all these magnitudes are uncorrelated, it results[43] in

$$u(k_1) = [u^2_{N_{if}}(k_1) + u^2_M(k_1) + u^2_{\Delta\varphi}(k_1)]^{1/2}, \tag{20}$$

where

$$u_{N_{if}}(k_1) = \left|\frac{\partial k_1}{\partial N_{if}}\right| u(N_{if}) = \frac{k_1}{N_{if}} u(N_{if}), \tag{21}$$

$$u_M(k_1) = \left|\frac{\partial k_1}{\partial M}\right| u(M) = \frac{k_1}{M} u(M), \tag{22}$$

$$u_{\Delta\varphi}(k_1) = \left|\frac{\partial k_1}{\partial \Delta\varphi_{3m}}\right| u(\Delta\varphi_{3m}) = \frac{M}{N_{if}} u(\Delta\varphi_{3m}) \tag{23}$$

are the uncertainties in k_1 due to N_{if}, M, and $\Delta\varphi_{3m}$, respectively.

As the distance N_{if} is approximately constant in all the measurements (we always consider points separated about 75% of the image width) and, as we will see later, M, $u(N_{\mathrm{if}})$, $u(M)$, and $u(\Delta\varphi_{3m})$ are also constant, the two first terms of $u(k_1)$ are proportional to k_1 and the third one is constant. On the other hand, if we employ as variable of interest the Lamb wavelength λ_1, its uncertainty $u(\lambda_1)$ can also be split in three terms, two of them proportional to λ_1 and the third one proportional to λ_1 squared. Therefore, the variation of the uncertainty with respect to the experimental point is smaller if we consider k_1 than if we consider λ_1 as experimental datum.

The distance N_{if}, in pixels, is measured in the digital image as the difference between the pixel number of two points, so $u(N_{if}) = \sqrt{2}u(N_j)$, being $u(N_j)$ the uncertainty in the position N_j of a point P_j given in pixels. The value of $u(N_j)$ was estimated as the standard deviation of a square distribution having as width 2 pixels. Then,

$$u(N_{if}) = \sqrt{2}\frac{1}{\sqrt{3}}. \tag{24}$$

For a typical value of $N_{if}=750$, $u_{N_{if}}(k_1) \approx 0.001 k_1$.

The magnification M is measured from the digital image of a scale located at the same position than the plate surface, i.e.,

$$M = \frac{N_L}{L}, \tag{25}$$

being N_L the number of pixels in the image and L the actual distance in the scale. Neglecting the uncertainty in the scale, the uncertainty in M is given by

$$u(M) = \frac{1}{L} u(N_L). \tag{26}$$

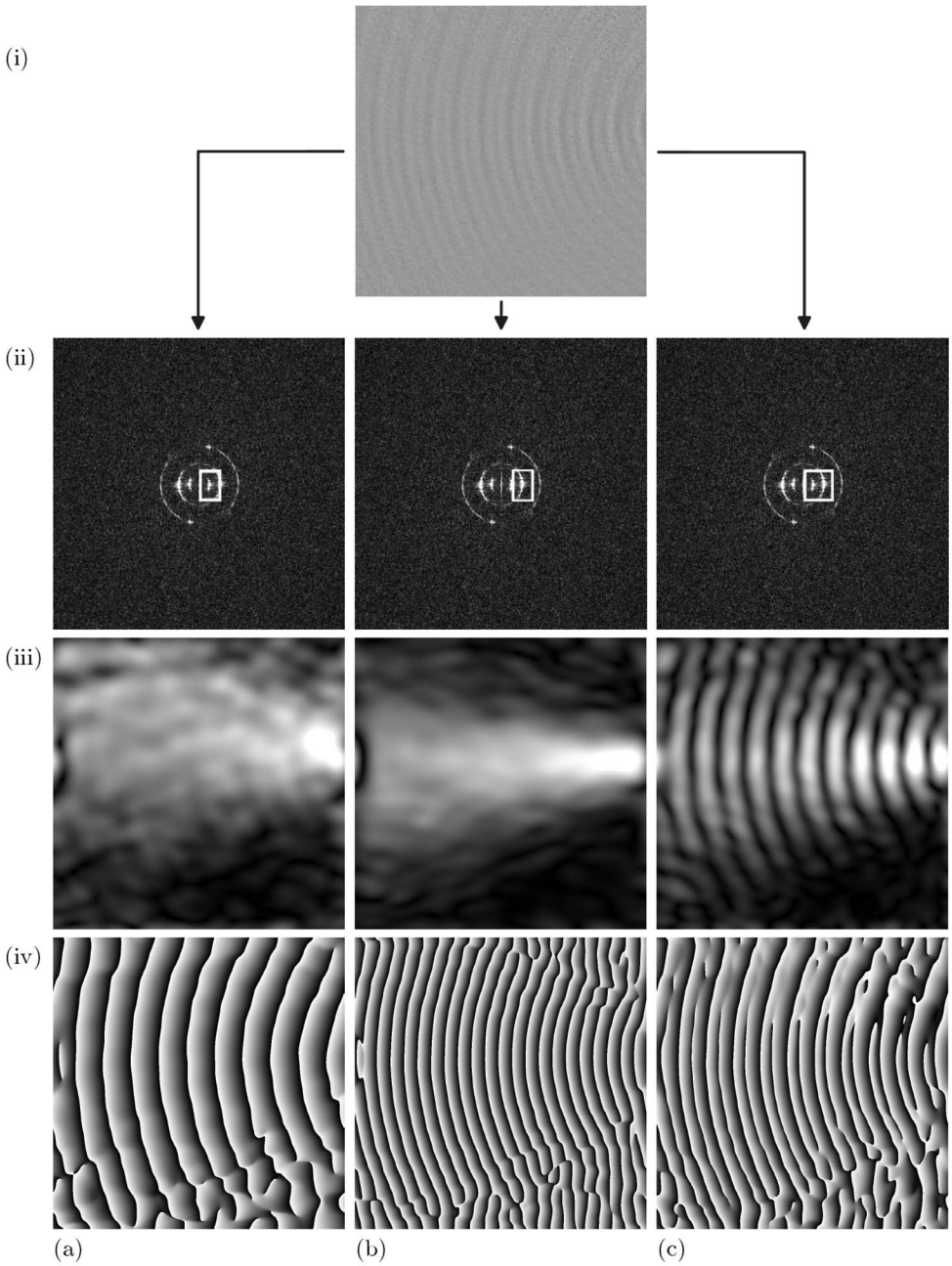


FIG. 7. Out-of-plane displacement map at a given instant t_1 of a Lamb wave consisting of modes $A1$ and $S0$ at 0.952 MHz in an aluminum plate 1.97 mm thick, measured with our double-pulsed ESPI system. The actual size of the field of view is 111×111 mm^2. (i) Instantaneous optical phase-change map, $\Delta\Phi(\mathbf{x},t_1)$. (ii) Fast Fourier transform of the optical phase-change map indicating the position of the filter. (iii) Acoustic amplitude, $u_{3m}(\mathbf{x})$. (iv) Acoustic phase, $\arg[\hat{u}_3(\mathbf{x},t_1)]$. (a) A filter is applied in order to isolate mode $A1$. (b) A filter is applied in order to isolate mode $S0$. (c) A coarser filter is applied so that both modes $A1$ and $S0$ are recovered and the beating between them is detected.

Therefore, considering again that the measured length is equivalent to 750 pixels and that $u(N_L)=u(N_{if})$, it results in $u_M(k_1)\approx 0.001k_1$.

On the other hand, the uncertainty in $\Delta\varphi_{3m}$ depends on several factors. Among them, we can highlight the quantization error due to the discretization of the output signal of each pixel in the charge coupled device camera, the speckle noise present in the optical phase-change maps, and the fact that the assumption of plain wave fronts propagating in the horizontal direction is not exact (Fig. 6). To estimate the value of $u_{\Delta\varphi}(k_1)$ we repeated several times the measurement of the same k_1 and calculated the standard deviation of the measured values. Each set of measurements for a given k_1 was performed with constant magnification and considering always the same single mode [like the one shown in Fig. 6(a)] at the same frequency (so that k_1 is constant). In this case, although the magnification is not an exact value and has an uncertainty, k_1 is obtained from Eq. (18) considering a fixed value of M, and then the standard deviation of the measurements is not affected by the uncertainty in M, i.e., it corresponds to

$$s(k_1) = [u_{N_{if}}^2(k_1) + u_{\Delta\varphi}^2(k_1)]^{1/2}. \tag{27}$$

It was found that $s(k_1)$ is approximately constant, about 10 m^{-1}, for all the modes, except for those having a value of

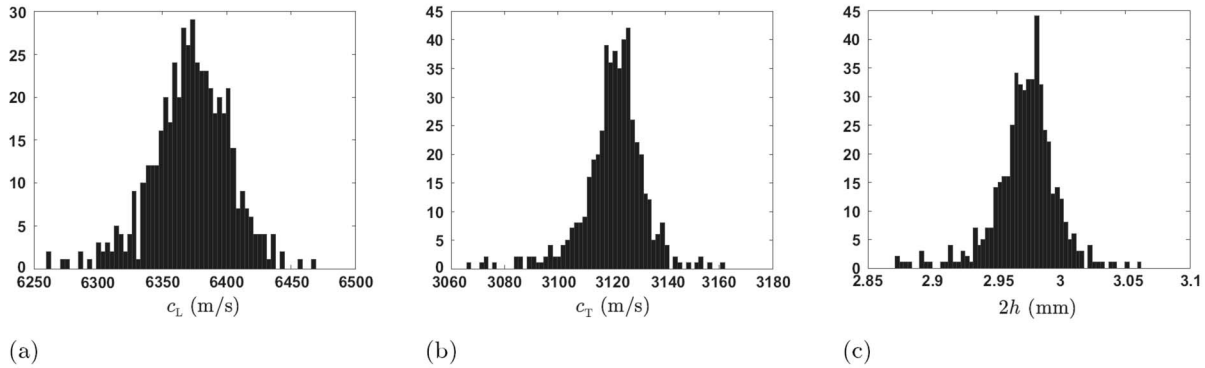


FIG. 8. Histograms showing the dispersion of the calculated values of the parameters by means of the Monte Carlo method for the plate 2.98 mm thick. (a) Longitudinal wave phase velocity c_L. (b) Shear wave phase velocity c_T. (c) Thickness of the plate $2h$.

$k_1 > 2000$ m^{-1}, for which a slight increment of the uncertainty is perceived as k_1 increases due to the term $u_{N_{if}}(k_1)$. Therefore, it can be assumed that $u_{\Delta\varphi}(k_1) \approx 10$ m^{-1}, which for a typical value of $N_{if}/M = 0.08$ m is equivalent to $u(\Delta\varphi_{3m}) \approx 0.8$ rad.

Another major source of uncertainty in $\Delta\varphi_{3m}$ has to do with the presence of one or more undesired modes. However, provided the amplitudes of these modes are small enough, the uncertainty induced in k_1 can be neglected against the uncertainty in k_1 due to the other sources (see the Appendix).

Therefore, the value of $u(k_1)$ is approximately equal to the constant factor $u_{\Delta\varphi}(k_1)$, except maybe for those experimental points having a high value of k_1 (over $k_1 = 2000$ m^{-1}), but in any case the variation is small. Furthermore, as shown in Fig. 9, there are few points in the region of high wave number, so they do not affect greatly the fitting. To perform the calculations we considered a constant value of $u(k_1) = 15$ m^{-1}.

The wave number of the modes could also be measured in the Fourier spectrum. However, the separation between two consecutive frequencies is $1/W$ (W being the width of the image), so the uncertainty in k_1 measured from the peaks in the spectrum is in the order of $2\pi/W$, i.e., approximately 62 m^{-1} for a typical value of $W = 0.1$ m.

IV. FITTING THE MODEL PARAMETERS TO THE EXPERIMENTAL DATA

The procedure we followed to obtain the model parameters c_L, c_T, and $2h$ consists in the minimization of the error function χ_{k_1} defined as[44]

$$\chi_{k_1}(c_L, c_T, 2h) = \left\{ \sum_i [k_{1i} - k_1(f_i, c_L, c_T, 2h)]^2 \right\}^{1/2}, \qquad (28)$$

being k_{1i} the experimental wave number and k_1 its theoretical value calculated with Eq. (1) from the experimental frequency f_i and the values $(c_L, c_T, 2h)$ of the parameters to be determined.

The sum of Eq. (28) is usually weighted with the inverse of the standard deviation of the experimental magnitude.[44] However, as we have already shown in Sec. III E, the standard deviation of the wave number is approximately constant regardless of the frequency of the experiment, so that the error function is simplified and the fitting procedure is easier. That is the reason why we used this magnitude instead of the directly measured wavelength.

The values of the parameters that minimize $\chi_{k_1}(c_L, c_T, 2h)$ correspond to the most likely solution of the plate thickness and bulk wave velocities in the plate material, and hence they are considered to be the best estimation of the values of such constants. To minimize the error function, we applied a downhill simplex algorithm described in Ref. 44 taking as initial guesses the values for the same samples calculated by a different technique described in Ref. 18.

The experimental frequencies f_i and wave numbers k_{1i} are assumed to present normal distributions with standard deviations $s(f_i) = 5$ kHz and $s(k_{1i}) = 15$ m^{-1}, respectively. Then, the uncertainties in the calculated parameters depend on the uncertainties in $2n$ experimental magnitudes, n being the number of experimental points (k_{1i}, f_i), that makes impractical the use of the so-called law of propagation of uncertainties. Therefore, to estimate the uncertainty in c_L, c_T, and $2h$ we applied the Monte Carlo method,[44] calculating these parameters 500 times for each plate by using sets of input quantities (k_{1i}, f_i), with $i = 1, 2, \ldots, n$, generated randomly according to the probability density function that we assigned. This method estimates the shape of the probability density function of the calculated parameters, permitting to assess nonlinear effects. Thereby, the uncertainties in the calculated parameters are equal to the standard deviation of the resulting set of values (that is, we have employed a coverage factor $k = 1$). An example of a histogram showing the dispersion of the results is displayed in Fig. 8 for the plate 2.98 mm thick. The calculated thickness and bulk wave velocities and their respective uncertainties are displayed in Table I, along with the values of the longitudinal phase velocity and the thickness measured by an independent method.

The uncertainties in $H = E/\rho$ and ν displayed in Table I are calculated as combined standard deviations[43] from the uncertainties in c_L and c_T. However, the problem can be aimed to find the values of H, ν, and $2h$ by minimizing the error function expressed as

$$\chi_{k_1}(H, \nu, 2h) = \left\{ \sum_i [k_{1i} - k_1(f_i, H, \nu, 2h)]^2 \right\}^{1/2}. \qquad (29)$$

The results of this alternative approach are shown in Table II together with the uncertainties obtained with the

TABLE II. Calculated values of the thickness $2h$, the constant H, and Poisson's ratio ν obtained from minimizing $\chi_{k_1}(H,\nu,2h)$ with n experimental points. The value of Young's modulus E is obtained as the product $H\rho$, being ρ the mass density.

$2h$ (mm) (outside micrometer)	n	$2h$ (mm)	H (10^7 m^2/s^2)	ν	ρ (kg/m^3)	E (10^9 N/m^2)
1.51 ± 0.01	34	1.511 ± 0.006	2.601 ± 0.008	0.3420 ± 0.0028	2790 ± 15	72.57 ± 0.46
1.97 ± 0.01	38	1.969 ± 0.008	2.596 ± 0.009	0.3423 ± 0.0025	2790 ± 15	72.43 ± 0.46
2.98 ± 0.01	46	2.976 ± 0.020	2.618 ± 0.016	0.3424 ± 0.0018	2790 ± 15	73.04 ± 0.59
4.03 ± 0.01	67	4.014 ± 0.022	2.595 ± 0.013	0.3441 ± 0.0015	2790 ± 15	72.40 ± 0.52
5.00 ± 0.01	70	4.997 ± 0.025	2.604 ± 0.011	0.3440 ± 0.0012	2790 ± 15	72.65 ± 0.50

Monte Carlo method. In general, the uncertainties in H and ν obtained directly with the Monte Carlo method are smaller than the ones obtained from the uncertainties in c_L and c_T, but in any case the difference is not significant.

Figure 9 shows the theoretical frequency spectra for the calculated parameters of each plate given in Table I, together with the set of experimental points. The fitting is rather good for all the plates.

V. CONCLUSIONS

A new application of a well-known remote detection, full-field optical technique, namely, pulsed ESPI, was described to characterize plates made of an isotropic and homogeneous material, aluminum in our case. Our self-developed variant of this technique allows a single-shot recording of the acoustic field, yielding a two dimensional map of the out-of-plane instantaneous displacements of transient Lamb waves propagating along the plate, so that it is possible to freeze the wave motion before it reaches any edge, avoiding disturbing reflection phenomena. Furthermore, a processing of the map allows one to evaluate the instantaneous acoustic amplitude and phase maps of the wave.

From those maps, the wavelength λ_{1i} of a given narrowband single-mode Lamb wave of known frequency f_i can be accurately measured. Repeating the process for other single modes at different frequencies, a set of points (λ_{1i},f_i) is obtained, from which the thickness of the plate as well as the bulk wave velocities (or, alternatively, two independent elastic constants) can be determined, provided enough measurements (at least, three) are taken.

A preliminary study was made in order to determine the best modes and frequencies to yield a precise solution. Then, a considerable number of experimental points were measured within the available frequency range. Finally, a fitting procedure based on the minimization of an error function by means of the downhill simplex algorithm was applied. The uncertainties in the calculated parameters c_L, c_T, and $2h$ were estimated with the Monte Carlo method resulting that they were smaller than 1%. Even more accurate results can be

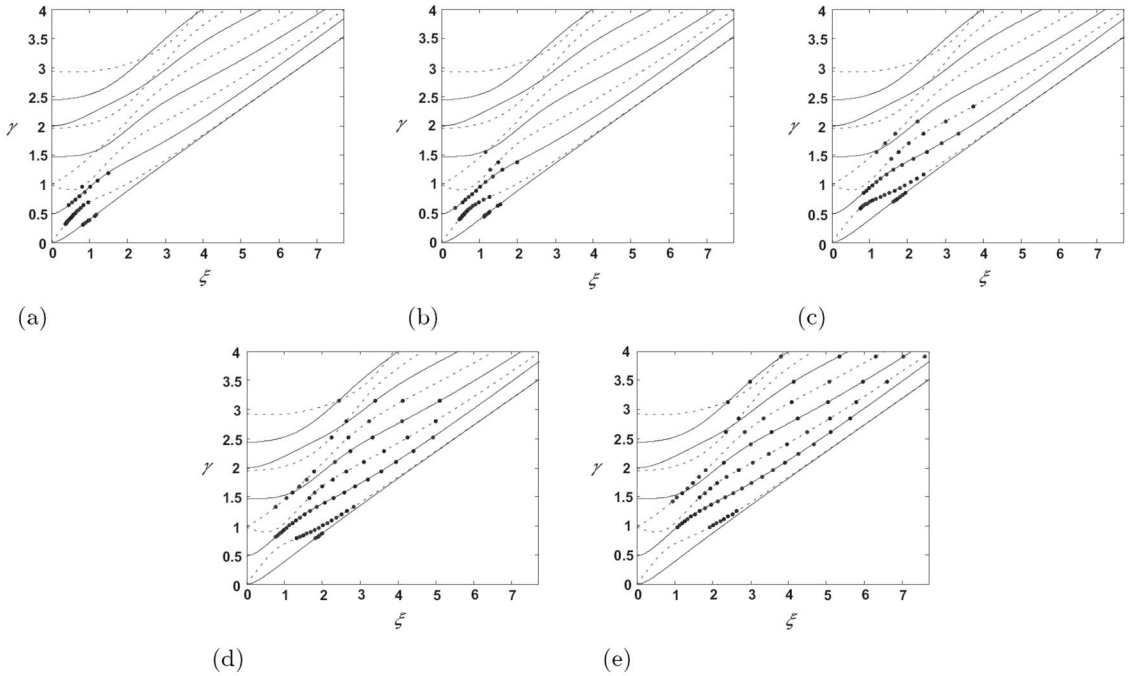


FIG. 9. Experimental points together with the theoretical representation of the frequency spectra of Lamb modes for the calculated thickness and elastic constants. For the identification of each mode, see Fig. 1. (a) Plate 1.51 mm thick. (b) Plate 1.97 mm thick. (c) Plate 2.98 mm thick. (d) Plate 4.03 mm thick. (e) Plate 5.00 mm thick.

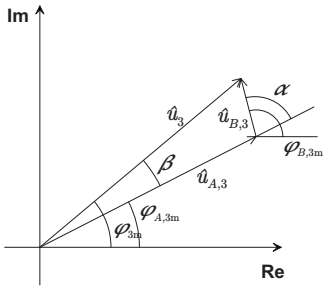


FIG. 10. Complex representation of the out-of-plane displacement at a point of the plate surface when there are two modes A and B propagating along the plate.

achieved if some of these parameters are precisely known and only one or two of them have to be fitted.

ACKNOWLEDGMENTS

This work was cofunded by the Spanish Ministerio de Educación y Ciencia and by the European Commission (ERDF) in the context of the Plan Nacional de I+D+i (Project No. DPI2005-09203-C03-01) and by the Dirección Xeral de Investigación, Desenvolvemento e Innovación da Xunta de Galicia in the context of the Plan Galego de IDIT (Project No. PGIDIT06PXIC303193PN). Supplementary cofunding from the Universidade de Vigo (Project No. I608122F64102) is also acknowledged. The authors would like to thank Professor Mariano Pérez-Amor for his advice and encouragement throughout the development of our double-pulsed ESPI system. The authors acknowledge the contribution of calibration samples from Tecnatom, S. A. and Applus Norcontrol, S.L.U. Machining of ultrasound generating hardware and especial optomechanic elements were made by the technician Mr. Pablo Barreiro.

APPENDIX: UNCERTAINTY IN k_1 DUE TO THE PRESENCE OF A SECOND MODE

When, apart from the desired mode A, there is a second mode B propagating along the plate, the instantaneous out-of-plane displacement field at $t=0$ is given by

$$u_3(\mathbf{x}) = \mathrm{Re}[u_{3m}(\mathbf{x})\exp\{j\varphi_{3m}(\mathbf{x})\}] = \mathrm{Re}[\hat{u}_{A,3}(\mathbf{x}) + \hat{u}_{B,3}(\mathbf{x})], \tag{A1}$$

where $\hat{u}_{A,3}(\mathbf{x})$ and $\hat{u}_{B,3}(\mathbf{x})$ are given by Eq. (6). Hence, as we can see in Fig. 10, there is a difference β between the phase we detect in the phase map $\varphi_{3m}(\mathbf{x})$ and the one we want to measure $\varphi_{A,3m}(\mathbf{x})$. The value of β can be calculated from the phase difference $\alpha=\varphi_{B,3m}-\varphi_{A,3m}$ between the two modes. When the amplitude of the second mode $u_{B,m}$ is much smaller than the amplitude of the first one $u_{A,m}$, β is given approximately by

$$\beta \approx \frac{u_{B,3m}}{u_{A,3m}} \sin\alpha. \tag{A2}$$

The uncertainty in $\varphi_{A,3m}$ can be estimated as the standard deviation $s(\beta)$. To calculate $s(\beta)$ we presumed, for simplicity, that the value of α at the measuring points P_i or P_f is completely random, i.e., it presents a rectangular distribution between $-\pi$ and π, and so its probability density function (pdf) corresponds to

$$f_\alpha(\alpha) = \frac{1}{2\pi}. \tag{A3}$$

The inverse relation

$$\alpha = \arcsin\frac{u_{A,3m}}{u_{B,3m}}\beta \tag{A4}$$

has two roots, one in the interval ($|\alpha|<\pi/2$) and the other in the interval ($\pi/2<|\alpha|<\pi$), and therefore the pdf $f_\beta(\beta)$ is calculated as

$$f_\beta(\beta) = \frac{f_\alpha[\alpha_1(\beta)]}{\left|\left(d\beta/d\alpha\right)_{\alpha_1}\right|} + \frac{f_\alpha[\alpha_2(\beta)]}{\left|\left(d\beta/d\alpha\right)_{\alpha_2}\right|}, \tag{A5}$$

where α_1 belongs to the first interval and α_2 to the second one. Hence, from Eqs. (A2) and (A3) it results in

$$f_\beta(\beta) = \frac{1}{\pi\sqrt{\left(u_{B,3m}/u_{A,3m}\right)^2 - \beta^2}} \tag{A6}$$

that takes nonzero values in the interval $(-u_{B,3m}/u_{A,3m}, u_{B,3m}/u_{A,3m})$ and has a mean value (expectation) of zero. The variance of the distribution described by Eq. (A6) is given by

$$s^2(\beta) = \int_{-u_{B,3m}/u_{A,3m}}^{u_{B,3m}/u_{A,3m}} \frac{\beta^2}{\pi\sqrt{\left(u_{B,3m}/u_{A,3m}\right)^2 - \beta^2}} d\beta \tag{A7}$$

and can be calculated making the variable change $\beta=|u_{B,3m}/u_{A,3m}|\sin\theta$, resulting in

$$s^2(\beta) = \int_{-\pi/2}^{\pi/2} \frac{1}{\pi}\left(\frac{u_{B,3m}}{u_{A,3m}}\right)^2 \sin^2\theta d\theta = \frac{1}{2}\left(\frac{u_{B,3m}}{u_{A,3m}}\right)^2. \tag{A8}$$

Then, the uncertainty in the acoustic phase due to the presence of a weak second mode can be estimated by

$$u(\varphi_{A,3m}) \approx s(\beta) = \frac{u_{B,3m}}{\sqrt{2}u_{A,3m}}. \tag{A9}$$

Taking $u_{A,3m}=5u_{B,3m}$ the uncertainty $u(\Delta\varphi_{A,3m})=\sqrt{2}u(\varphi_{A,3m})=0.2$ rad, and for a typical value of $N_{if}/M=0.08$ m the resulting uncertainty in k_1 is $u(k_1)=2.5\ \mathrm{m}^{-1}$, i.e., negligible against other sources. We checked this result with several measurements taken under repeatability conditions in which a soft beating is detected [like the one shown in Fig. 6(b)], resulting in a value of $s(k_1)$ approximately equal to the one obtained when there is only one appreciable mode.

[1] *Nondestructive Testing Handbook*, edited by A. S. Birks, R. E. Green, and P. McIntire (American Society for Nondestructive Testing, Columbus, OH, 1991).
[2] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1975).
[3] I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[4] C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications* (Hilger, Bristol, 1990), pp. 148–222.
[5] T. D. Mast and G. A. Gordon, "Quantitative flaw reconstruction from ultrasonic surface wavefields measured by electronic speckle pattern interferometry," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 432–444 (2001).
[6] J. L. Fernández, A. F. Doval, C. Trillo, J. L. Deán, and J. C. López, "Video

ultrasonics by pulsed TV holography: A new capability for non-destructive testing of shell structures," International Journal of Optomechatronics **1**, 122–153 (2007).
[7]R. J. Dewhurst, C. Edwards, A. D. W. McKie, and S. B. Palmer, "Estimation of the thickness of thin metal sheet using laser generated ultrasound," Appl. Phys. Lett. **51**, 1066–1068 (1987).
[8]R. Zhang, M. Wan, and W. Cao, "Parameter measurement of thin elastic layers using low-frequency multi-mode ultrasonic Lamb waves," IEEE Trans. Instrum. Meas. **50**, 1397–1403 (2001).
[9]B. Culshaw, G. Pierce, and P. Jun, "Non-contact measurement of the mechanical properties of materials using an all-optical technique." IEEE Sens. J. **3**, 62–70 (2003).
[10]W. Gao, C. Glorieux, and J. Thoen, "Laser ultrasonic study of Lamb waves: Determination of the thickness and velocities of a thin plate," Int. J. Eng. Sci. **41**, 219–228 (2003).
[11]M. Lematre, Y. Benmehrez, G. Bourse, J. W. Xu, and M. Ourak, "Determination of elastic parameters in isotropic plates by using acoustic microscopy measurements and an optimization method," NDT & E Int. **35**, 493–502 (2002).
[12]A. S. Murfin and R. J. Dewhurst, "Estimation of wall thinning in mild steel using laser ultrasound Lamb waves and a non-steady-state photo-emf detector," Ultrasonics **40**, 777–781 (2002).
[13]X. Chen and M. Wan, "Parameter measurement of the cylindrically curved thin layer using low-frequency circumferential Lamb waves," Ultrasonics **43**, 357–364 (2005).
[14]Y. H. Kim, S.-J. Song, S.-D. Kwon, Y.-M. Cheong, and H.-K. Jung, "Determination of ultrasonic wave velocities and phase velocity dispersion curves of an Inconel 600 plate using resonant ultrasound spectroscopy and leaky Lamb waves," Ultrasonics **42**, 551–555 (2004).
[15]D. Clorennec, C. Prada, and D. Royer, "Local and noncontact measurements of bulk acoustic wave velocities in thin isotropic plates and shells using zero group velocity Lamb modes," J. Appl. Phys. **101**, 034908 (2007).
[16]W. P. Rogers, "Elastic property measurement using Rayleigh-Lamb waves," Res. Nondestruct. Eval. **6**, 185–208 (1995).
[17]A. Safaeinili, O. I. Lobkis, and D. E. Chimenti, "Quantitative materials characterization using air-coupled leaky Lamb waves," Ultrasonics **34**, 393–396 (1996).
[18]J. L. Fernández, J. L. Deán, C. Trillo, and A. F. Doval, "Elastic constants determination by direct measurement of the beat wavelength between *A*0 and *S*0 Lamb modes with pulsed TV holography," Opt. Lasers Eng. **45**, 618–630 (2007).
[19]Y. Hayashi, S. Ogawa, H. Cho, and M. Takemoto, "Non-contact estimation of thickness and elastic properties of metallic foils by the wavelet transform of laser-generated Lamb waves," NDT & E Int. **32**, 21–27 (1999).
[20]S. E. Bobbin, J. W. Wagner, and R. C. Cammarata, "Determination of the flexural modulus of thin films from measurement of the first arrival of the symmetric Lamb wave," Appl. Phys. Lett. **59**, 1544–1546 (1991).
[21]J. A. Rogers, L. Dhar, and K. A. Nelson, "Noncontact determination of transverse isotropic elastic moduli in polyimide thin films using a laser based ultrasonic method," Appl. Phys. Lett. **65**, 312–314 (1994).
[22]T.-T. Wu and Y.-H. Liu, "On the measurement of anisotropic elastic constants of fiber-reinforced composite plate using ultrasonic bulk wave and laser generated Lamb wave," Ultrasonics **37**, 405–412 (1999).
[23]Y. Mizutani and M. Takemoto, "A simplified elastic stiffness estimation of unidirectional carbon-fiber-reinforced coupon using the in-plane velocity anisotropy of Lamb waves," Jpn. J. Appl. Phys., Part 1 **37**, 3110–3115 (1998).
[24]C. M. Hernandez, T. W. Murray, and S. Krishnaswamy, "Laser ultrasonic measurement of mechanical properties of nanometer-sized thin film structures," Res. Nondestruct. Eval. **21**, 1187–1192 (2002).
[25]Y.-C. Lee and S.-W. Cheng, "Measuring Lamb wave dispersion curves of a bi-layered plate and its application on material characterization of coating," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 830–837 (2001).
[26]O. Behrend, A. Kulik, and G. Gremaud, "Characterization of thin films using numerical inversion of the generalized Lamb wave dispersion relation," Appl. Phys. Lett. **62**, 2787–2789 (1993).
[27]M. J. Banet, M. Fuchs, J. A. Rogers, J. H. Reinhold, Jr., J. M. Knecht, M. Rothschild, R. Logan, A. A. Maznev, and K. A. Nelson, "High-precision film thickness determination using a laser-based ultrasonic technique," Appl. Phys. Lett. **73**, 169–171 (1998).
[28]M. R. Karim, A. K. Mal, and Y. Bar-Cohen, "Inversion of leaky Lamb wave data by simplex algorithm," J. Acoust. Soc. Am. **88**, 482–491 (1990).
[29]J. L. Deán, J. L. Fernández, A. F. Doval, and C. Trillo, "Measurement of Lamb waves dispersion curves under narrowband monomode excitation using TV holography," Proc. SPIE **6341**, 63410B (2006).
[30]S. Kenderian, B. B. Djordjevic, and R. E. Green, "Narrow band laser-generated surface acoustic waves using a formed source in the ablative regime," J. Acoust. Soc. Am. **113**, 261–266 (2003).
[31]C. Edwards, A. C. Bushell, S. B. Palmer, and H. Nakano, "Laser interference grating for surface wave generation," Nondestr. Test. Eval. **10**, 15–23 (1992).
[32]J. Huang, S. Krishnaswamy, and J. D. Achenbach, "Laser generation of narrow-band surface waves," J. Acoust. Soc. Am. **92**, 2527–2531 (1992).
[33]K. Yamanaka, O. V. Kolosov, Y. Nagata, T. Koda, H. Nishino, and Y. Tsukahara, "Analysis of excitation and coherent amplitude enhancement of surface acoustic waves by the phase velocity scanning method," J. Appl. Phys. **74**, 6511–6522 (1993).
[34]J. A. Rogers and K. A. Nelson, "Study of Lamb acoustic waveguide modes in unsupported polyimide thin films using real-time impulsive stimulated thermal scattering," J. Appl. Phys. **75**, 1534–1556 (1994).
[35]K. V. de Rostyne, C. Glorieux, W. Gao, W. Lauriks, and J. Thoen, "Experimental investigation of leaky Lamb modes by an optically induced grating," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1245–1253 (2002).
[36]J. L. Rose, *Ultrasonic Waves in Elastic Media* (Cambridge University Press, Cambridge, 2000), pp. 200–240.
[37]V. K. Kinra and V. R. Iyer, "Ultrasonic measurement of the thickness, phase velocity, density or attenuation of a thin-viscoelastic plate. Part I: the forward problem," Ultrasonics **33**, 95–109 (1995).
[38]V. K. Kinra and V. R. Iyer, "Ultrasonic measurement of the thickness, phase velocity, density or attenuation of a thin-viscoelastic plate. Part II: the inverse problem," Ultrasonics **33**, 111–122 (1995).
[39]M. Wan, B. Jiang, and W. Cao, "Direct measurement of ultrasonic velocity of thin elastic layers," J. Acoust. Soc. Am. **101**, 626–628 (1997).
[40]A. F. Doval, "A systematic approach to TV holography," Meas. Sci. Technol. **11**, R1-R36 (2000).
[41]C. Trillo, A. F. Doval, D. Cernadas, O. López, J. C. López, B. V. Dorrío, J. L. Fernández, and M. Pérez-Amor, "Measurement of the complex amplitude of transient surface acoustic waves using double-pulsed TV holography and a two-stage spatial Fourier transform method," Meas. Sci. Technol. **14**, 2127–2134 (2003).
[42]C. Trillo, D. Cernadas, A. F. Doval, C. López, B. V. Dorrío, and J. L. Fernández, "Detection of transient surface acoustic waves of nanometric amplitude with double-pulsed TV holography," Appl. Opt. **42**, 1228–1235 (2003).
[43]"ISO, Guide to the expression of uncertainty in measurement," Geneva (1993).
[44]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1988).

Experimental study of transient paths to the extinction in sonoluminescence

Raúl Urteaga, Damián Dellavale, Gabriela F. Puente, and Fabián J. Bonetto
Laboratorio de Cavitación y Biotecnología, Instituto Balseiro-CAB-CONICET, R8402AGP–Av. Bustillo Km. 9,5–San Carlos de Bariloche-Río Negro, Argentina

(Received 5 June 2007; revised 2 January 2008; accepted 6 March 2008)

An experimental study of the extinction threshold of single bubble sonoluminescence in an air-water system is presented. Different runs from 5% to 100% of air concentrations were performed at room pressure and temperature. The intensity of sonoluminescence (SL) and time of collapse (t_c) with respect to the driving were measured while the acoustic pressure was linearly increased from the onset of SL until the bubble extinction. The experimental data were compared with theoretical predictions for shape and position instability thresholds. It was found that the extinction of the bubble is determined by different mechanisms depending on the air concentration. For concentrations greater than ~30%–40% with respect to the saturation, the parametric instability limits the maximum value of R_0 that can be reached. On the other hand, for lower concentrations, the extinction appears as a limitation in the time of collapse. Two different mechanisms emerge in this range, i.e., the Bjerknes force and the Rayleigh–Taylor instability. The bubble acoustic emission produces backreaction on the bubble itself. This effect occurs in both mechanisms and is essential for the correct prediction of the extinction threshold in the case of low air dissolved concentration. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2903854]

PACS number(s): 43.35.Hl [AJS] Pages: 1490–1496

I. INTRODUCTION

Sonoluminescence is the phenomenon where a pulsed light source is obtained from an ultrasonic field. Given that the light belongs to the visible spectra, a concentration of 12 orders of magnitude in energy density is obtained.[1] This phenomenon occurs when a bubble is trapped in an antinode pressure (velocity node) of a standing wave. When the pressure amplitude is increased, the bubble oscillations become highly nonlinear and produce a violent collapse. During the collapse, an almost adiabatic heating of the inner gas is achieved, leading to the thermal emission process.[1,2]

The region of the parameter space in which this phenomenon occurs is limited by two main factors. First, the collapse must be strong enough to produce the necessary heating of the bubble contents. This implies that the point in the phase diagram must be above the Blake threshold.[3] Second, the bubble must remain stable during the whole process. In addition to the diffusive and chemical stabilities, which determine the equilibrium position in the phase space, the shape and position stability must be satisfied, i.e., the bubble must not break into pieces and it has to remain trapped in the pressure antinode.

Holt and Gaitan[4] and Gaitan and Holt[5] studied the stability boundaries of single bubble sonoluminescence (SBSL). They found that two upper limits appear. One limit is in the ambient radius (R_0) and the other is in the acoustic pressure (P_a) in which stable SBSL is obtained.

The upper threshold in R_0 has been interpreted as the onset of the parametrical instability, in reasonable accordance with the experimental data.[5–8] This limit extends to the region where the bubble is not sonoluminescent. However, regarding the limit in P_a, there is a controversy as to which mechanism is responsible for the extinction.[9]

In addition, the experimental data collected in the region of the SBSL parameter space by most groups[4–7] have been obtained in static conditions of acoustic pressure. In this way, the points in the vicinity of extinction are inaccessible because they are, in general, unstable.

The instability in the vicinity of the extinction threshold is associated with a recycling mode.[4,5,10,11] Furthermore, it has been divided into two types depending on the temporal pattern. Type 1 is associated with a sudden drop in intensity and time of collapse, followed by a slow increase of both quantities in the order of seconds. Type 2 is similar to type 1, but the maximum intensity is tenfold smaller and has a colder spectrum than stable sonoluminescence (SL).[10] Those modes of recycling are related to a bubble pinch off and formation of multiple fragments. These fragments can dissolve or recombine to form a new bubble with a renewed intake of air. In type 1 recycling, the nitrogen and oxygen are burned off while the equilibrium diffusive concentration is reached. In type 2, instead, the instability threshold is reached before the burn off becomes efficient and, therefore, the bubble is mostly air filled.[10]

Dam and Levinsen[10] observed that both types of recycling modes are preceded by a period doubling in the amplitude of SL. This fact and the spatial anisotropy in the flashes are indications that the instability would be instigated by a nonspherical perturbation. In addition, An *et al.*[9] calculated the instability threshold of SL under a nonspherical symmetrical acoustic pressure perturbation. They chose this perturbation as a parameter to fit the available experimental data without specifying the source of such a perturbation.

Besides, Holzfuss and Holt[12] showed that the shock wave emitted by the bubble generates a complex mode field consisting of higher harmonics of the applied driving. These

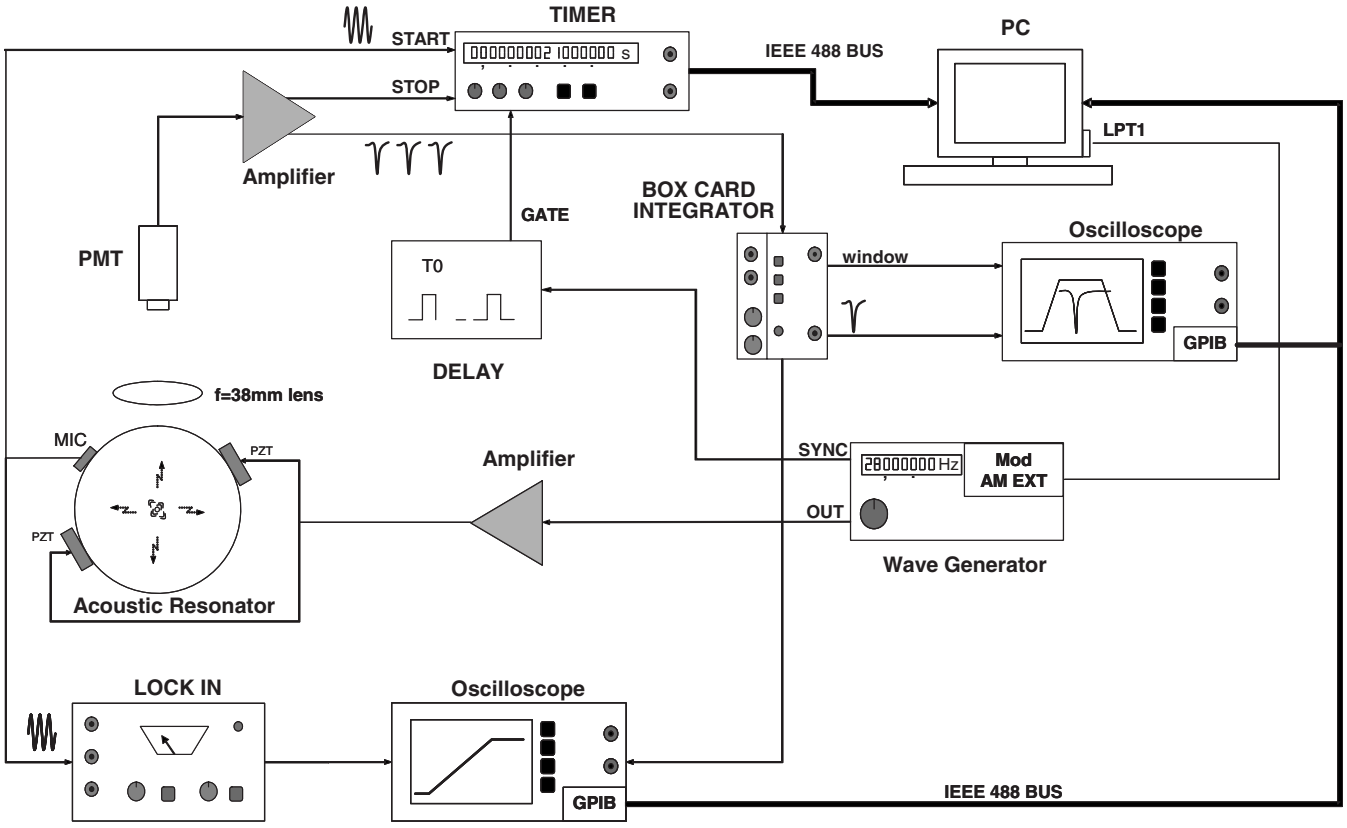


FIG. 1. Experimental setup. Typical apparatus for SBSL in which a driving voltage ramp is applied. The SL intensity, microphone signal, and time of collapse are acquired as a function of time.

harmonics extend to the megahertz range and have about 10% of the driving amplitude. Holzfuss and Holt found that the bubble position and stability are determined by this complex sound field.

In this work, we measure the data points in phase space in a continuous way by slowly increasing the acoustic pressure. As a consequence, we can access the region close to the extinction. Repetitions of these experiments give us an indication of the statistical occurrence of the extinction threshold. Moreover, we can explore the region in which the dynamics are unstable and obtain an estimation of the strength of instability.

The results are compared with theoretical predictions of the extinction threshold for different instabilities of shape and position. A close relationship between the extinction and the high frequency harmonics emitted by the bubble is proposed.

II. APPARATUS

Measurements were made by using a typical apparatus for SBSL similar to the one first reported in Ref. 8. We use a Pyrex made spherical resonator 60 mm in outer diameter with a resonance frequency of approximately 28.3 kHz. The resonator cell was filled with distilled, 0.45 μm filtered, and partially degassed water. After filling, the liquid sample was sealed to prevent diffusion to the atmosphere. The experiments were conducted at a controlled temperature of 23 °C and a room pressure of 0.92 bar. The whole setup was thermally isolated and temperature controlled. We surveyed the phase space of SBSL by making a continuous increase in the voltage of the drivers from the onset of SL until the extinction of the bubble for a set of different concentrations of dissolved air. The air concentrations were between 5% and 100% with respect to the saturation. To determine the position in the phase space at each moment, we simultaneously measured the time of collapse (defined as the time between the zero crossing of acoustic pressure with negative slope and the bubble collapse), the SL flash intensity, and the microphone signal.

Figure 1 shows the experimental setup. A function generator signal was amplified by an audio amplifier and a resonant *RLC* circuit. In such a circuit, the capacitance was the piezodrivers (PZT) of the resonator. The output of the function generator was computer controlled to generate the amplitude-modulated wave for the resonator driving.

In each run, the frequency was adjusted to the resonant frequency of the system. The resonant frequency changes indicate the slight increase in temperature because of the piezodriver heating.

We linearly increased the acoustic pressure from the onset of SL, at about 1.2 bar, until the bubble extinction at about 1.4 bar. To choose the acoustic pressure increase rate, two facts must be taken into account. First, the rate of increase must be small enough to ensure a negligible departure from the behavior under a constant acoustic pressure at each instant. On the other hand, the rate of the increase must be as high as possible to explore the unstable extinction region. In

order to satisfy these constraints, we choose approximately 4 mbar/s for the acoustic pressure increase rate (i.e., each run takes about 50 s).

The pressure values at the beginning and end slightly depend on the particular dissolved air concentration. In all the runs, the pressure increase rate was the same.

We measured the SL pulse intensity using an Oriel 33740 phototube (PT). We obtained the bubble time of collapse and the SL intensity from the PT signal. A small piezoelectric crystal (microphone) was glued to the resonator wall to obtain a signal proportional to the applied acoustic pressure. The time of collapse was measured with a Stanford Research timer (SR620). The timer start and stop signals were obtained from the microphone and PT pulse, respectively. The time precision of the measurements was only limited by the microphone signal jitter (about 50 ns). The timer was operated at its maximum rate of about 1200 samples/s, which is close to $\frac{1}{24}$ of the resonator frequency.

The microphone signal was filtered by a lock-in amplifier referenced to the function generator output. Thus, a signal proportional to the resonator driving amplitude was obtained. The integration time of the lock-in was 10 ms in all the measurements.

A gated integrator boxcar Stanford Research Systems (SRS) averages the intensity of 1000 samples of the SL flash at a frequency of about 14 kHz (half of the driving frequency). The lock-in output and the SL intensity were acquired by a digital oscilloscope connected via General Purpose Instrumentation Bus (GPIB) to the computer.

Another digital oscilloscope was used to monitor the integrator window (3 μs in width) and the SL pulse coincidence in time. The SL pulse was preamplified 25 times by using a Stanford Research amplifier (SR244). Hence, the PC acquired the time of collapse, the microphone, and the SL intensity as a function of the measurement time.

The temporal evolution of the bubble radius was measured using the Mie scattering technique (not shown in Fig. 1)[8,13–15] at different acoustic pressure values for each air dissolved concentration. In these measurements, the values of acoustic pressure were constant in time.

We use the numerical model described in Ref. 8 to fit the bubble radius temporal evolution. We outline here the different components of our model. See Ref. 8 for more details. The model includes the following: (1) transient and spatially nonuniform heat transfer using a collocation point method, dissociation of O_2 and N_2, and mass diffusion of vapor in the noncondensable gases inside the bubble; (2) nonequilibrium evaporation and condensation of water and a temperature jump due to the accommodation coefficient at the bubble interface; and (3) transient and spatially nonuniform heat transfer using a collocation point method and mass diffusion of the gas in the liquid. The model computes the gas diffusion in the liquid and inside the bubble following a similar approach to the one used for the thermal equation by obtaining the points in phase space that are stable from a mass diffusive point of view.[16] The chemical kinetic model describes the molecular decomposition of the water vapor molecules inside the bubble upon collision with an inert gas. It includes the elementary reactions between stable species and free radicals for the system.

The acoustic pressure obtained from the model was used to calibrate the filtered microphone signal. A calibration of the microphone was made for each dissolved air concentration. The calibrations do not show statistical variation for the different air concentrations.

The ambient radius (R_0) was computed by the numerical model[8] from the acoustic pressure and time of collapse values. We take R_0 at the time when the acoustic pressure is zero with negative slope. We want to remark that, while it is well established that the determination of acoustic pressure with this technique is quite good (0.05 bar precision),[1,8,15] the value of R_0, instead, will be correctly predicted only in case a chemical and diffusive equilibrium inside the bubble is developed.

Finally, it is important to note that because of the measurement system, based on the detection of the flash pulses, the parameter space obtained is strictly for sonoluminescence.

III. RESULTS

Figures 2(a)–2(c) show the time of collapse, the SL intensity, and the microphone signal, respectively, as functions of the measurement time in an experimental run. The air concentration was about 5% with respect to saturation. We set the bubble extinction time as $t=0$ s.

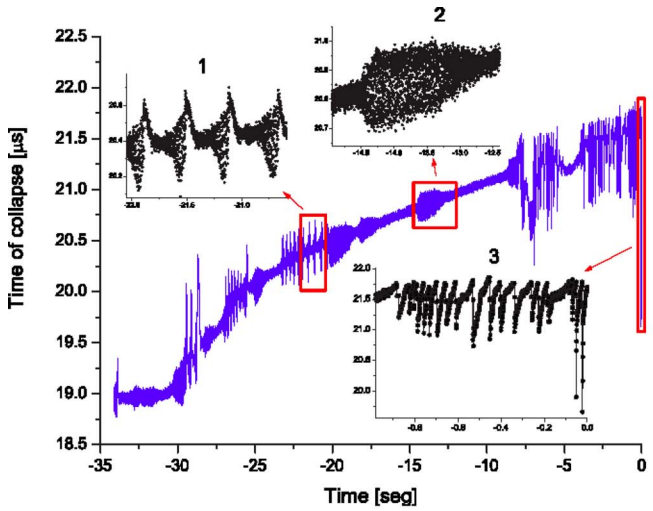
In this particular experiment, the ramp was started at $t=-30$ s, increasing the driving signal amplitude until the bubble extinction at $t=0$ s.

The time of collapse, the SL intensity, and the microphone signal amplitude mean values monotonically increase. However, several fluctuations appear in the whole range, as is shown in Fig. 2. At a time of about −8 s, a different behavior arises with particular features until the extinction of the bubble.

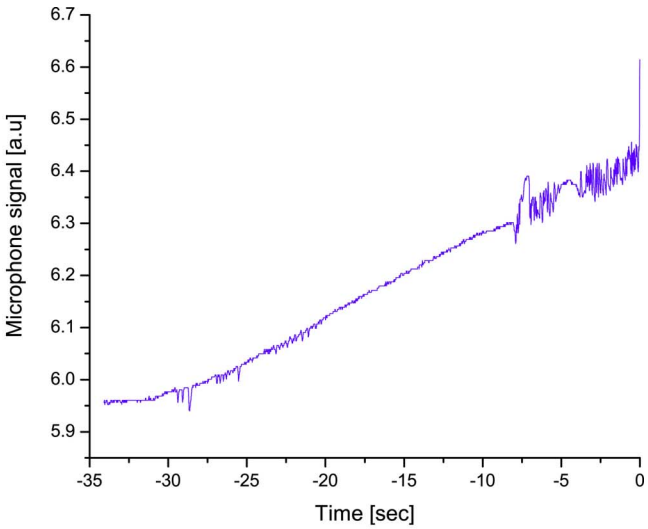
The bursts observed in Fig. 2(a) at times $t=-20$ s and $t=-14$ s (details in the insets) were common in all the measurements regardless of the air concentration. The particular shape and temporal position of the burst were different for each measurement. However, for a given air concentration, the ramp zones where the time of collapse has low variance are quite repetitive. This fact will be further analyzed from the results of Fig. 3.

Figure 2(a), insets 1 and 2, show that the time of collapse fluctuations are quite complex. In some cases (inset 2), the structure is similar to the period doubling reported in the past.[17] Time of collapse fluctuations as those displayed by inset 1 have not been reported in the literature.

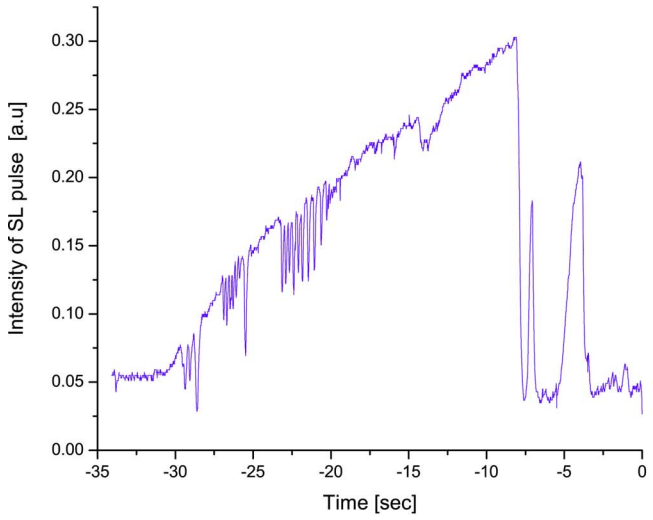
The SL intensity [Fig. 2(b)] decreases in coincidence with the time of collapse burst. At higher acoustic pressures and before the bubble extinction, a series of sudden drops in time of collapse appears. At the same time, a tenfold decrease in SL intensity is observed. This behavior is similar to the one investigated by Dam and Levinsen[10] and named the second recycling mode. In Ref. 10, a period doubling is found just before that recycling mode appears. We are not



(a)



(c)



(b)

FIG. 2. (Color online) (a) Time of collapse temporal evaluation for 5% dissolved air with respect to saturation. The insets show details of fluctuations and the extinction. (b) Light pulse intensity for the same run as in (a). The intensity results from a running average of 1000 pulses. (c) Driving frequency component of the microphone signal.

able to detect this feature because of the running average performed by the boxcar integrator in the SL intensity determination.

Two stability regions can be recognized from Fig. 2, at times $t=-7$ s and $t=-4$ s. In those regions, the time of collapse is almost constant and the SL intensity slowly increases. Note that the slow increase is not an integration effect but a diffusive process within the bubble. The drops, instead, seem to be limited only by the SL pulse integration.

Figure 2, inset 3, shows an enlargement of the final stages of the bubble before the extinction. In such a region, the occurrence frequency of recycling is quite high (30–100 Hz). In the last stage, the amplitude of recycling also increases and the bubble suddenly disappears.

Figure 3 shows the SL intensity as a function of measurement time for six realizations at ~5% of air concentration with respect to saturation. The SL intensity monotonically increases with the time of collapse almost until the bubble extinction.

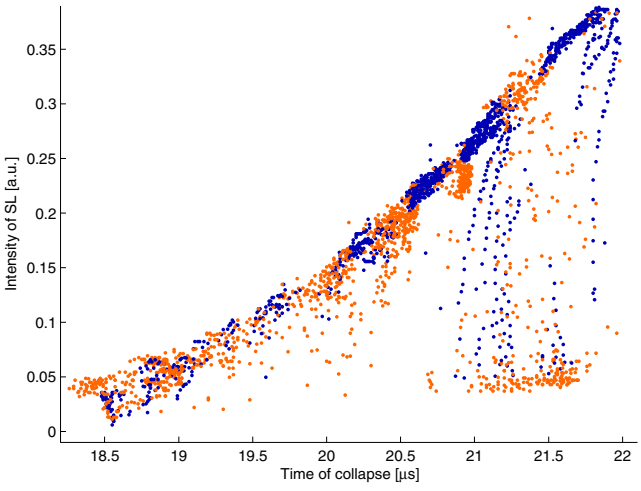


FIG. 3. (Color online) The SL intensity measured during the ramp in voltage of PZT for six runs at ~5% of air concentration with respect to saturation as a function of time of collapse. Stable (black dots, blue online) and unstable (gray dots, red online) points are separated. It is considered stable if the maximum difference at the time of collapse of 500 consecutive points is lower than 150 ns.

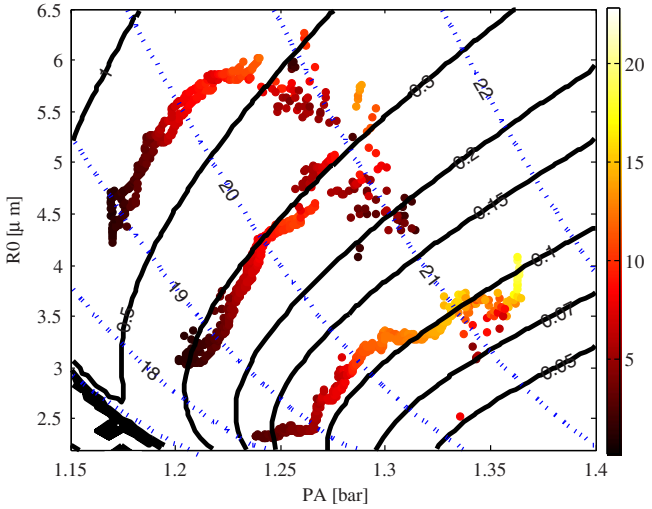


FIG. 4. (Color online) Phase diagram of sonoluminescence. The points are the experimental results for three different air concentrations (from bottom to top: 0.14, 0.4, and 0.5, respectively). For each air concentration, two runs are shown. The gray level (color online) maps the SL intensity relative to the scale at the right. The filled black lines are the equilibrium curves for different air concentrations. The dashed (blue online) lines are the constant time of collapse curves.

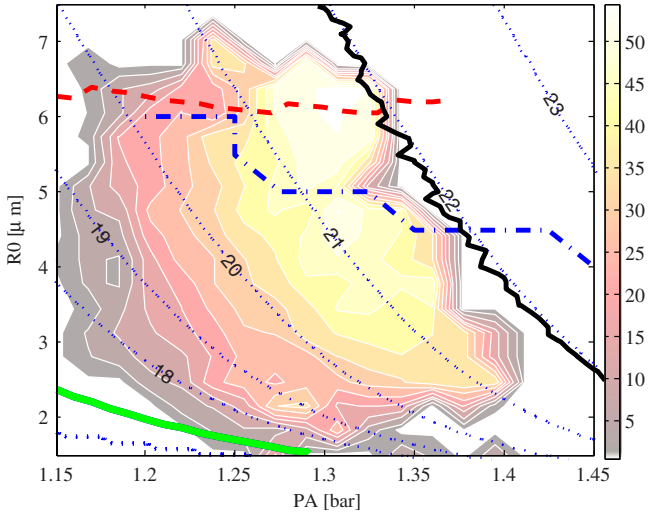


FIG. 5. (Color online) Phase space of sonoluminescence. The thin dashed (blue online) lines are the constant time of collapse curves. The thick dashed (red online) line is the parametric instability for the mode $n=2$ (Refs. 1 and 8); the dotted dashed (blue online) line is the Rayleigh–Taylor instability for the mode $n=2$ (Refs. 1 and 8); the filled gray (green online) line is the Blake threshold. The filled black line is the Bjerknes force threshold where a high frequency harmonic of the acoustic pressure has been taken into account in the calculation.

However, before the bubble extinction, the SL intensity presents a decrease until values as low as 10% of the maximum SL intensity. The decrease in SL intensity is coincident with the second recycling mode in the last stage of Fig. 2.

In Fig. 3, "stable" and "unstable" points are classified. This differentiation was obtained by computing the maximum time of collapse difference among 500 consecutive points. When that difference was bigger than 150 ns, the points were considered unstable; otherwise, they were considered stable.

Figure 3 shows separated zones for stable and unstable points especially at high acoustic pressure values. Also, the stability islands of Fig. 2 correspond to stable point region of Fig. 3. The ambient radius (R_0) and acoustic pressure (P_a) were computed from the measured time of collapse and microphone signal, respectively.

Figure 4 displays the R_0 and P_a values for three different air concentrations (the measured values of C_{inf}/C_0 were 0.14, 0.4, and 0.5). To illustrate the measurement fluctuations, two runs are shown for each air concentration. The SL intensity is mapped into the color bar at the right of Fig. 4.

The assigned errors of the points are 50 ns for the time of collapse (t_c) and 0.05 bar for the acoustic pressure (P_a). The acoustic pressure errors at high air dissolved concentrations are larger (0.1 bar) due to the relatively poor bubble spatial stability. The ambient radius error is determined by the P_a and t_c errors and the position in the phase diagram. The R_0 error is bounded to 0.7 μm for small R_0 values and 1.5 μm for high R_0 values.

As Fig. 4 shows, the SL intensity monotonically increases with the acoustic pressure for the three air concentrations. Nevertheless, a region of low SL intensity for times of collapse at about 21 μs can be observed. This region corresponds to the low SL intensity points at the right of Fig. 3. Figure 4 shows that the measured values of air concentration reasonably agree with the ones predicted by the numerical model. While the pathways for $C_{\text{inf}}/C_0=0.4$ and 0.5 follow the trends of the equilibrium curves, the path for $C_{\text{inf}}/C_0=0.14$ shows a more pronounced deviation at low acoustic pressure.

For measured air concentrations higher than 60%, the points lie on stability curves of lower calculated air concentrations. For measured air concentrations of 82% and 100%, the points were always unstable and spread over the whole phase diagram around the time of collapse curve of 21 μs.

Figure 5 displays R_0 and P_a values for 11 air concentrations in the range of 5%–100%. About ten runs in each concentration were measured. To adequately represent the high number of data points, the SL intensity was averaged into patch regions of 0.08 μm and 6 mbar over the whole phase diagram. The SL intensity is mapped into the color bar at the right of the figure.

Figure 5 shows that the SL intensity increases with the acoustic pressure for all air concentrations. Besides, it is worth noting that the SL intensity seems to be determined only by the time of collapse value, regardless of the air concentration.

In Fig 5, other than the stable equilibrium and time of collapse curves, the calculated parametric instability, Rayleigh–Taylor instability, and Bjerknes instability threshold are also included. In order to close the phase space of SL, the Blake threshold is included. The Bjerknes instability was calculated by taking into account two acoustic field contributions: The first component was the acoustic field generated by the drivers at the frequency of resonance (f_{res}) and the second contribution was an 11-fold in frequency and a 10% in amplitude with regard to the first contribution. The relative phase between them was set to zero. At $t=0$ the fundamental crosses zero with negative slope ($-\sin(2\pi f_{\text{res}}t)$). We define

the relative phase equal to zero when at $t=0$ the harmonic (second contribution of the acoustic field) crosses zero with negative slope too. Similar results were obtained by considering another harmonic. On the other hand, the phase between the harmonics is a sensitive parameter. We chose the phase in order to obtain the calculated extinction threshold for the smallest time of collapse value.

With regard to the extinction threshold, it can be observed that for calculated concentrations higher than 30%, the extinction of the bubble is determined by an upper value in equilibrium radius. This fact is in good accordance with the limit set by the parametric instability.

For calculated air concentrations lower than 30%–40%, we found that the evolution of SL presents an upper limit in the time of collapse between 21.5 and 22 μs.

The Rayleigh–Taylor instability calculated as in Ref. 8 is not consistent with the threshold established by the experimental data.

IV. DISCUSSION

From the onset of SL at low acoustic pressure, the light emission process is stable except for the existence of zones in which burst in time of collapse appears [Fig. 2(a), insets 1 and 2]. The positions of those zones are reproducible in different runs for a fixed air dissolved concentration and temperature. These bursts present a complex structure. In some cases, a period doubling and even more sophisticated evolutions were observed.

With the exception of some earlier works,[5,12] these interesting structures have been largely overlooked in SL. We believe that it is because the bursts are experimentally obtained only in small regions and present a high sensitivity upon the control parameters such as temperature or dissolved gases.

At higher acoustic pressures and before the extinction, a recycling mechanism is developed [Fig. 2(a), inset 3]. It is possible to recognize two types of recycling modes that differ in the recycling frequency and SL intensity. In Fig. 2(a), a zone of the first mode is developed between regions of the second recycling mode at the end of the temporal evolution.

The first recycling mode was predominant at high air concentrations and appears at lower acoustic pressure values than the second type. We found that the extinction of the bubble always occurs after the second recycling mode is developed, regardless of the air concentration (Fig. 4). The measurement time (i.e., the range of acoustic pressures) that the bubble persists in the second recycling mode increases with the dissolved air concentration; as a matter of fact, for air saturated water, the bubble is in a recycling mode in the whole range of acoustic pressures where SL is observed.

We interpret the extinction as a statistical event in which each pinch off of the bubble in the recycling mode is a potential extinction. We experimentally found that the extinction probability is a function mostly of the time of collapse of the bubble. This is true for air concentration lower than 30%–40%. For higher concentrations, the parametric instability seems to be the mechanism that produces the extinction.

Taking into account the characteristics of the recycling mode, specifically, the existence of a period doubling in SL amplitude and correlated spatial anisotropy,[11] and the pinch off of the bubble in fragments,[10,11] we associated the onset of this phenomenon to a nonspherical acoustic pressure perturbation.

We perform a simple analysis of the intensity of ultrasound emitted by the bubble from the temporal evolution of the radius (from the numerical model).[3] The results show that the total emission of ultrasound also mainly depends on the time of collapse. This preliminary calculation is not conclusive but gives us an indication that a possible source for the perturbation is actually the ultrasound emission of the bubble that interacts with the resonator. Furthermore, there is evidence[12,15,18,19] that the ultrasound emission of the bubble affects the position and stability of the bubble.

In the calculated instability threshold of SL by An *et al.*[9] a nonspherical symmetrical acoustic pressure perturbation is utilized as a parameter to fit the available experimental data without specifying the source of such a perturbation. However, the calculated threshold shows the correct dependence at low air dissolved concentrations.

On the other hand, we present the results on the Bjerknes instability when an acoustic harmonic is considered. This calculation does not pretend to evaluate the exact value of the Bjerknes force from the complete ultrasonic field that experiences the bubble but evaluates the reliability of the mechanism. The stability threshold obtained fits rather well the experimental data. Nevertheless, it is worth noting that the stability threshold does not imply the bubble extinction. In addition, if the Bjerknes force is the mechanism that limits the values of the time of collapse, it is expected that the SL intensity smoothly varies while the bubble goes away from the center of the resonator. This feature was never observed in the experiments.

V. CONCLUSIONS

We study the evolution of the SL while a continuous increase in acoustic pressure was applied. We found two different behaviors determined by the air dissolved concentration. For concentration higher than 40%–50% with respect to saturation, the extinction of the bubble is determined by an upper value in equilibrium radius. This fact is in good agreement with the limit set by the parametric instability.

For air dissolved concentrations lower than 30%–40%, we found that the evolution of SL presents a different mechanism for the extinction. We experimentally found that the extinction probability is a function mostly of the time of collapse value.

We propose two different mechanisms that could explain this behavior: A Rayleigh–Taylor instability amplified by a nonspherical acoustic perturbation and the existence of a destabilizing Bjerknes force that takes into account the contribution of the acoustic harmonics.

Although both mechanisms show the correct dependence, the mechanism of the nonspherical acoustic perturbation is more consistent with the experimental observation of an abrupt extinction of the bubble. However, more calcula-

tions including the measured value of the perturbations are necessary to validate the proposed mechanism. Nevertheless, in both mechanisms analyzed, the backreaction of the bubble acoustic emissions seems to be essential for the correct prediction of the extinction threshold for low air concentrations.

ACKNOWLEDGMENTS

The authors thank Daniel Mateos, Sebastián Eckardt, and Enrique Aburto for technical support. This work was funded by CONICET and CNEA.

[1]M. P. Brenner, S. Hilgenfeldt, and D. Lohse, "Single-bubble sonoluminescence," Rev. Mod. Phys. **74**, 425–484 (2002).
[2]Y. An and C. F. Ying, "Model of single bubble sonoluminescence," Phys. Rev. E **71**, 036308 1–12 (2005).
[3]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[4]R. G. Holt and D. F. Gaitan, "Observation of stability boundaries in the parameter space of single bubble sonoluminescence," Phys. Rev. Lett. **77**, 3791–3794 (1996).
[5]D. F. Gaitan and R. G. Holt, "Experimental observations of bubble response and light intensity near the threshold for single bubble sonoluminescence in an air-water system," Phys. Rev. E **59**, 5495–5502 (1999).
[6]J. A. Ketterling and R. E. Apfel, "Extensive experimental mapping of sonoluminescence parameter space," Phys. Rev. E **61**, 3832–3837 (2000).
[7]J. A. Ketterling and R. E. Apfel, "Shape and extinction thresholds in sonoluminescence parameter space," J. Acoust. Soc. Am. **107**, L13–L18 (2000).
[8]G. F. Puente, R. Urteaga, and F. J. Bonetto, "Numerical and experimental study of dissociation in an air-water single-bubble sonoluminescence system," Phys. Rev. E **72**, 046305 1–10 (2005).
[9]Y. An, T. Lu, and B. Yang, "Instability of sonoluminescing bubbles under a nonspherical symmetrical acoustic-pressure perturbation," Phys. Rev. E **71**, 026310 (2005).
[10]J. S. Dam and M. T. Levinsen, "Second mode of recycling together with period doubling links single-bubble and multibubble sonoluminescence," Phys. Rev. Lett. **94**, 174301 (2005).
[11]J. S. Dam, M. T. Levinsen, and M. Skogstad, "Stable nonspherical bubble collapse including period doubling in sonoluminescence," Phys. Rev. E **67**, 026303 (2003).
[12]J. Holzfuss and R. G. Holt, "Acoustical stability of a sonoluminescing bubble," Phys. Rev. E **66**, 046630 (2002).
[13]B. P. Barber and S. J. Putterman, "Light scattering measurements of the repetitive supersonic implosion of a sonoluminescing bubble," Phys. Rev. Lett. **69**, 3839–3842 (1992).
[14]B. Gompf and R. Pecha, "Mie scattering from a sonoluminescing bubble with high spatial and temporal resolution," Phys. Rev. E **61**, 5253–5256 (2000).
[15]T. J. Matula, S. M. Cordry, R. A. Roy, and L. A. Crum, "Bjerknes force and bubble levitation under single-bubble sonoluminescence conditions," J. Acoust. Soc. Am. **102**, 1522–1527 (1997).
[16]S. Hilgenfeldt, D. Lohse, and M. Brenner, "Phase diagrams for sonoluminescing bubbles," Phys. Fluids **8**, 2808–2826 (1996).
[17]R. G. Holt, D. F. Gaitan, A. A. Atchley, and J. Holzfuss, "Chaotic sonoluminescence," Phys. Rev. Lett. **72**, 1376–1379 (1994).
[18]J. Holzfuss, M. Ruggeberg, and A. Billo, "Shock wave emissions of a sonoluminescing bubble," Phys. Rev. Lett. **81**, 5434–5437 (1998).
[19]T. J. Matula, "Inertial cavitation and single-bubble sonoluminescence," Philos. Trans. R. Soc. London, Ser. A **357**, 225–249 (1999).

On the forces in single-ended and push-pull electret transducers

Tim Mellow
Nokia UK Ltd., Farnborough, Hants GU14 0NG, United Kingdom

Leo Kärkkäinen
Nokia Research Center, 00180, Helsinki, Finland

(Received 14 January 2008; revised 20 May 2008; accepted 3 June 2008)

The equations for the electromechanical force conversion in single-ended and push-pull electret transducers are derived. Traditionally, the charge distribution has been modeled as a concentrated layer at an arbitrary distance from the surface of the dielectric. For the purpose of this analysis, a negative charge is assumed to be evenly distributed throughout the dielectric. The membrane has a conductive coating in which a positive charge is induced, giving an overall dipole charge. The resulting formulas are used to derive the voltage sensitivity of a microphone and the equivalent electrical circuit for the electromechanical transduction part of a microphone or loudspeaker. An equivalent external polarizing voltage is then derived that would produce the same driving force in a conventional electrostatic loudspeaker without a stored charge. The condition for the static stability of a circular electret membrane is also determined.
 [DOI: 10.1121/1.2951601]

PACS number(s): 43.38.Ar, 43.38.Bs, 43.20.Wd [AJZ] Pages: 1497–1504

I. INTRODUCTION

Since the electret microphone was introduced,[1,2] it has almost completely replaced all other types in the field of communications due to its robustness, reliability, and near studio quality. However, the application of an electret loudspeaker has remained largely elusive and electrostatic loudspeakers still invariably use an external polarizing supply in order to create a monopole charge in the conductive layer of the central membrane. The advantage of a monopole charge is that for a given membrane charge, the field does not vary with electrode separation. Electret membranes usually have a conductive coating which, while increasing the charge storage stability, produces a dipole charge. That is, a charge of the opposite polarity to that contained by the membrane is induced in the metallic coating. As a result, for a given membrane charge, the field decreases with increasing electrode separation. This is not such a major issue in the case of a microphone, where the electrode separation is typically very small, but in the case of a loudspeaker, wider separation is needed in order to achieve the volume acceleration that, in turn, is required to give a reasonable sound pressure level at some distance from it. Hence, membranes with a monopole charge have been investigated.[3,4] In recent years, electret membrane technology has been developing in terms of both the charge storage stability and maximum charge density. Porous membranes,[5,6] in particular, have received much attention and store enough charge to compensate for the dipole effect, having internal potentials of up to 500–1000 V. Hence, only dipole membranes are considered in this paper.

In accordance with Paschen's curve,[7] dipole charges are induced in the pores (pore size greater than 1 μm) by breaking down air inside the pores during the charging process. Both dipole and monocharges can be trapped in a porous electret dielectric. The pore size and pore density influences the relative density of dipole and monocharges. In general, the piezoelectric effect is found in porous electrets when the pore size is a few micrometers. On the other hand, reducing the size and density of the pores increases the amount of monocharge. In this paper, it is assumed that the monocharge dominates in a nanoporous material and, although the exact charge distribution is usually unknown, a possible distribution is shown in Fig. 1. It is likely that some charge will be lost at the outer faces, especially the one adjoining the conductive coating, where there will be some recombination with the positive charge induced in the coating. Hence, there will be a peak somewhere near the middle. However, in order to approximate this, a uniform charge distribution is assumed throughout the membrane.

For loudspeakers, another issue is that of space. Most pressure (or monopole) microphones convert pressure into mechanical displacement which in turn produces electrical charge displacement. Hence, they operate in the displacement controlled frequency range below the fundamental mechanical resonance. Loudspeakers, on the other hand, have to produce volume acceleration. Hence, they operate in the acceleration controlled frequency range above the fundamental mechanical resonance. Membrane loudspeakers have very little moving mass. In fact, the mass of the acoustic radiation load is usually greater than that of the membrane. In order to achieve a low enough fundamental resonance with such a small mass requires high mechanical compliance. Hence, a small enclosure is not an option. However, recent technological developments, such as activated carbon,[8,9] are moving in the right direction. Alternatively, a membrane loudspeaker can be used without an enclosure (i.e., as a dipole), but this requires a large diaphragm area in order to minimize rear-wave cancellation at low frequencies.

The theory of constant-charge push-pull electrostatic transducers with external polarizing supplies has long been established,[10] but there is little if any literature on the theory of push-pull electret transducers, although the force conver-

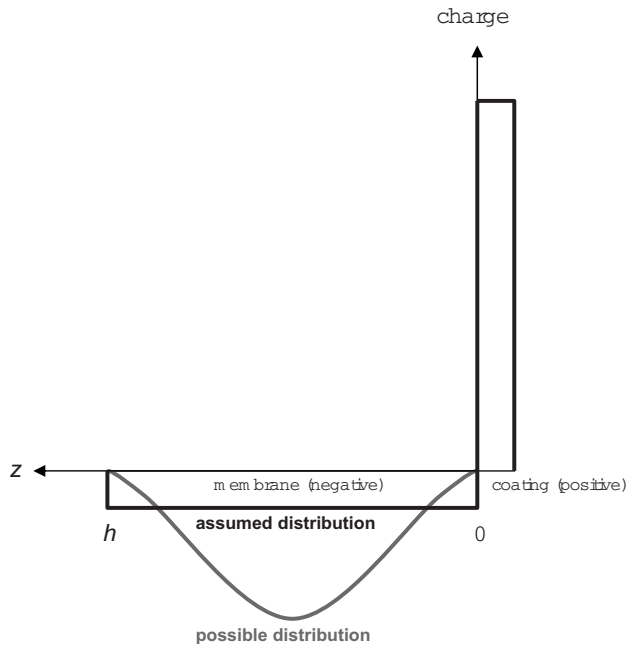


FIG. 1. Assumed charge distribution in membrane.

sion in single-ended electret transducer due to Sessler[11] is reproduced here by degenerating a push-pull model. Two push-pull configurations are analyzed in this paper. The first has a floating membrane, with just leakage paths between the conductive membrane coatings and the outer electrodes, represented by high-value resistors. Under dynamic conditions, these are ignored so that the configuration is treated as fully floating. However, under static conditions, the membrane is effectively grounded. Hence, the grounded membrane forms the basis of the second configuration to be analyzed. The forces and charges are also derived for a single-ended configuration, as well as a method to evaluate the stored charge density. An equivalent electrical circuit is developed which is intended as a basis for simulation, since sound radiation from membranes has already been analyzed in depth.[12–15] Finally, the condition for the static stability of a circular electret transducer is determined and compared with that of Streng[12] for a nonelectret transducer. The criteria for the stability of capacitance microphones have also been studied numerically by Warren,[16] although no closed-form conditional equations for specific geometries were presented.

II. PUSH-PULL ELECTRET TRANSDUCER DESCRIPTION

A push-pull configuration with a semifloating membrane is shown in Fig. 2. The two resistors R have very high values or could just represent leakage paths. Under quiescent conditions, with zero input signal, the membrane behaves as though it is strapped to the outer electrodes and hence the static charges assume the same values as in the case of a grounded membrane. This is the most stable steady state condition and also applies to very low frequencies, where half the input voltage is effectively connected between the membrane's conductive coatings and each outer electrode. However, when driven by a signal voltage across the outer electrodes at medium to high frequencies, the membrane behaves

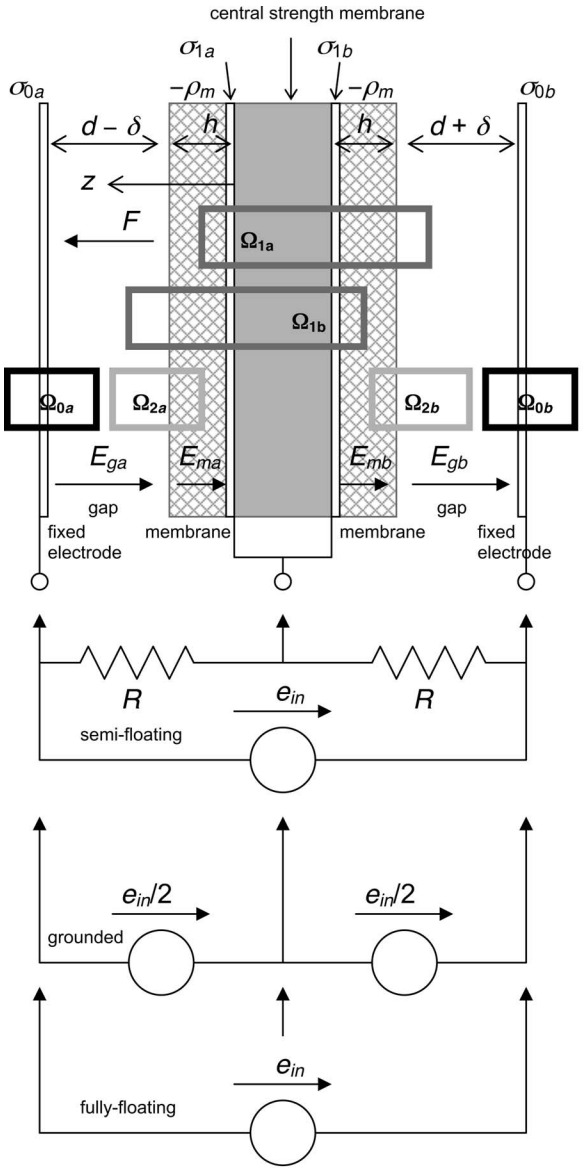


FIG. 2. Push-pull configuration with semifloating, grounded, and fully-floating membrane.

as though its conductive coatings are isolated and fully floating. This has a significant effect on the linearity of the membrane displacement.

The configuration in Fig. 2 is similar to one previously proposed by Atoji and Aoi,[17] except that their configuration has just one single charged membrane. Here, two are provided in order to maintain symmetry and to prevent the membrane from undergoing a permanent static displacement to one side, which would limit the maximum excursion. E_{ga} and E_{gb} are the electric field strengths in each of the gaps either side of the membrane, which actually comprises two membranes, each attached to a central strength membrane. Since the central strength membrane contains no charge, it does not contribute to the electric field and is thus assumed to be infinitesimally thin in the following force derivation. E_{ma} and E_{mb} are the electric field strengths within the two membranes, respectively, which are both assumed to contain an evenly distributed negative charge per unit volume (or volume charge density) $-\rho_m$. The inner electrode layers have

surface charge densities σ_{1a} and σ_{1b}. The outer plates have surface charge densities σ_{0a} and σ_{0b}. The relative permittivities of the membrane material and acoustic medium in the gaps are ε_r and ε_{r1}, respectively. If the gaps contain air, then ε_{r1} can be assumed to be unity.

III. GAUSS'S LAW

Gauss's law[18] states that the difference in electrical flux is equal to the net charge inside the volume. Applying this to the cylinders Ω_{0a} and Ω_{0b} gives

$$\varepsilon_0\varepsilon_{r1}E_{ga} = \sigma_{0a}, \tag{1}$$

$$-\varepsilon_0\varepsilon_{r1}E_{gb} = \sigma_{0b}. \tag{2}$$

It is assumed that the size of the electrodes is very large in comparison to the distance between them so that the field is perpendicular to the electrodes and the sides of the cylinders give zero flux. The tops of the cylinders are at constant distances from the outer electrodes, also the values of the electric fields are constant on the surfaces. Hence the fluxes are trivial to compute. Outside of the system, the field is assumed to be zero. The fluxes through the cylinders Ω_{2a} and Ω_{2b} are a little more complicated to calculate, since the membrane has a constant charge density:

$$\varepsilon_0\varepsilon_r E_{ma}(z) - \varepsilon_0\varepsilon_{r1}E_{ga} = -\rho_m(h-z), \tag{3}$$

$$\varepsilon_0\varepsilon_{r1}E_{gb} - \varepsilon_0\varepsilon_r E_{mb}(z) = -\rho_m(h+z). \tag{4}$$

Substituting Eq. (1) into Eq. (3) yields

$$E_{ma}(z) = \frac{\sigma_{0a} - (1 - z/h)\sigma_m}{\varepsilon_0\varepsilon_r} \tag{5}$$

and likewise substituting Eq. (2) into Eq. (4) yields

$$E_{mb}(z) = \frac{-\sigma_{0b} + (1 + z/h)\sigma_m}{\varepsilon_0\varepsilon_r}, \tag{6}$$

where $\sigma_m = \rho_m h$ is the constant charge per unit area (or surface charge density) of the membrane. Similarly, the fluxes through the cylinders Ω_{1a} and Ω_{1b} can be calculated as follows:

$$\varepsilon_0\varepsilon_{r1}E_{gb} - \varepsilon_0\varepsilon_r E_{ma}(z) = \sigma_{1a} + \sigma_{1b} - \rho_m(z+h), \tag{7}$$

$$\varepsilon_0\varepsilon_r E_{mb}(z) - \varepsilon_0\varepsilon_{r1}E_{ga} = \sigma_{1a} + \sigma_{1b} + \rho_m(z-h). \tag{8}$$

Substituting Eqs. (2) and (5) into Eq. (7) yields the following relation between the surface charge densities:

$$\sigma_{0a} + \sigma_{0b} + \sigma_{1a} + \sigma_{1b} - 2\sigma_m = 0. \tag{9}$$

Likewise, substituting Eqs. (1) and (6) into Eq. (8) yields the same relation between the surface charge densities

$$\sigma_{0a} + \sigma_{0b} + \sigma_{1a} + \sigma_{1b} - 2\sigma_m = 0. \tag{10}$$

Equations (1), (2), (5), and (6) give the functional forms of the electric fields, where E_{ga} and E_{gb} are constants and E_{ma} and E_{mb} are linear functions of z. The unknowns here are the surface charge densities σ_{0a}, σ_{0b}, σ_{1a}, and σ_{1b}. Their values depend on how the electrodes have been connected to each other.

IV. KIRCHHOFF'S LOOP RULE

A. Floating membrane

Now, let a potential (voltage) e_{in} be connected between the outermost electrodes, as shown in Fig. 2. In the fully floating state, the values of the resistors R are assumed to be infinite. This represents dynamic conditions at medium to high frequencies, where the impedance due to the interelectrode capacitance is much smaller than the impedance of the resistors. Using the loop equation (the direction of the electric fields is from left to right, but the integration from right to left) gives

$$\begin{aligned} -e_{in} &= -\int_{-h-d-\delta}^{-h} E_{gb}dz - \int_{-h}^{0} E_{mb}(z)dz - \int_{0}^{h} E_{ma}(z)dz \\ &\quad - \int_{h}^{h+d-\delta} E_{ga}dz \\ &= \frac{(\varepsilon_r(d+\delta) + \varepsilon_{r1}h)\sigma_{0b} - (\varepsilon_r(d-\delta) + \varepsilon_{r1}h)\sigma_{0a}}{\varepsilon_0\varepsilon_r\varepsilon_{r1}}. \end{aligned} \tag{11}$$

where the electric fields E_{gb}, E_{mb}, E_{ma}, and E_{ga} are given by Eqs. (2), (6), (5), and (1), respectively.

B. Grounded membrane

At very low frequencies, the impedance of the resistors R in Fig. 2 is much smaller than that of the interelectrode capacitance. Hence, half of the input voltage is connected across each resistor

$$\begin{aligned} -\frac{e_{in}}{2} &= -\int_{-h-d-\delta}^{-h} E_{gb}dz - \int_{-h}^{0} E_{mb}(z)dz \\ &= \frac{2(\varepsilon_r(d+\delta) + \varepsilon_{r1}h)\sigma_{0b} - \varepsilon_{r1}h\sigma_m}{2\varepsilon_0\varepsilon_r\varepsilon_{r1}}. \end{aligned} \tag{12}$$

$$\begin{aligned} -\frac{e_{in}}{2} &= -\int_{0}^{h} E_{ma}(z)dz - \int_{h}^{h+d-\delta} E_{ga}dz = \\ &\quad - \frac{(\varepsilon_r(d-\delta) + \varepsilon_{r1}h)\sigma_{0a} - \varepsilon_{r1}h\sigma_m}{2\varepsilon_0\varepsilon_r\varepsilon_{r1}}. \end{aligned} \tag{13}$$

where the electric fields E_{gb}, E_{mb}, E_{ma}, and E_{ga} are given by Eqs. (2), (6), (5), and (1) respectively.

V. SENSITIVITY AS A MICROPHONE

The voltage sensitivity is found by differentiating Eq. (11) with respect to displacement δ,

$$\Delta e_{in} = -\frac{\sigma_{0a} + \sigma_{0a}}{\varepsilon_0\varepsilon_{r1}}\Delta\delta. \tag{14}$$

For quiescent conditions ($\delta=0$, $e_{in}=0$), Eqs. (12) and (13) give

$$\sigma_{0a} = \sigma_{0b} = \frac{\varepsilon_{r1} h \sigma_m}{2(\varepsilon_r d + \varepsilon_{r1} h)}. \tag{15}$$

Hence, the linearized voltage sensitivity versus displacement of a floating push-pull microphone is defined by

$$\Delta e_{in} = -\frac{h\sigma_m}{\varepsilon_0(\varepsilon_r d + \varepsilon_{r1} h)} \Delta\delta. \tag{16}$$

The voltage sensitivity of a single-ended microphone is just half this value.

VI. CALCULATION OF THE NET FORCE AS A FUNCTION OF ELECTRODE CHARGES

The total energy W of the electric field can be calculated by integrating the dot product of the flux D and electric field E over the volume V [where $V=2(d+h)S$ and S is the surface area of the membrane] as follows:

$$W = \frac{1}{2}\int_V \bar{D}\cdot\bar{E} dV = \frac{S}{2}\int \bar{D}\cdot\bar{E} dz, \tag{17}$$

where

$$\bar{D} = \begin{cases} \varepsilon_0\varepsilon_{r1}\bar{E}_{gb}, & -h-d-\delta \le z < -h \\ \varepsilon_0\varepsilon_r\bar{E}_{mb}, & -h \le z \le 0 \\ \varepsilon_0\varepsilon_r\bar{E}_{ma}, & 0 \le z \le h \\ \varepsilon_0\varepsilon_{r1}\bar{E}_{ga}, & h \le z < h+d-\delta. \end{cases} \tag{18}$$

Hence,

$$W = \frac{S}{2}\varepsilon_0\left(\varepsilon_{r1}\int_{-h-d-\delta}^{-h} E_{gb}^2 dz + \varepsilon_r\int_{-h}^{0} E_{mb}^2 dz + \varepsilon_r\int_0^h E_{ma}^2 dz + \varepsilon_{r1}\int_h^{h+d-\delta} E_{ga}^2 dz\right), \tag{19}$$

where the electric fields E_{gb}, E_{mb}, E_{ma}, and E_{ga} are given by Eqs. (2), (6), (5), and (1), respectively. In this case, the energy density of the system will depend on δ, and there will be a force proportional to the charge of the membrane. The net force F acting on the membrane is then given by

$$F = \frac{\partial W}{\partial\delta} = \frac{S(\sigma_{0b}^2 - \sigma_{0a}^2)}{2\varepsilon_0\varepsilon_{r1}}. \tag{20}$$

VII. CHARGE DENSITIES

A. Floating membrane

Solving Eqs. (9) and (11) simultaneously for the charge densities σ_{0a} and σ_{0b} yields

$$\sigma_{0a} = \frac{(\varepsilon_r(d+\delta) + \varepsilon_{r1}h)(2\sigma_m - (\sigma_{1a}+\sigma_{1b})) + \varepsilon_0\varepsilon_r\varepsilon_{r1}e_{in}}{2(\varepsilon_r d + \varepsilon_{r1}h)}, \tag{21}$$

$$\sigma_{0b} = \frac{(\varepsilon_r(d-\delta) + \varepsilon_{r1}h)(2\sigma_m - (\sigma_{1a}+\sigma_{1b})) - \varepsilon_0\varepsilon_r\varepsilon_{r1}e_{in}}{2(\varepsilon_r d + \varepsilon_{r1}h)}. \tag{22}$$

It can be seen that the charge on the outer electrodes varies as the membrane moves from its central position, due to the charge movement between them via the voltage source. However, their sum is equal to the total charge density on the membrane as follows:

$$\sigma_{0a} + \sigma_{0b} = 2\sigma_m - (\sigma_{1a} + \sigma_{1b}) \tag{23}$$

and this remains constant. There may be similar charge movements between the charges σ_{1a} and σ_{1b} on the conductive coatings of the membranes, but this does not matter because it is the total membrane charge density that governs Eqs. (21) and (22).

B. Grounded membrane

Rearranging the loop rule of Eq. (13) yields an expressions for the charge density σ_{0a} as follows:

$$\sigma_{0a} = \frac{\varepsilon_{r1}(h\sigma_m + \varepsilon_0\varepsilon_r e_{in})}{2(\varepsilon_r(d-\delta) + \varepsilon_{r1}h)}. \tag{24}$$

Rearranging the loop rule of Eq. (12) yields an expressions for the charge density σ_{0b} as follows:

$$\sigma_{0b} = \frac{\varepsilon_{r1}(h\sigma_m - \varepsilon_0\varepsilon_r e_{in})}{2(\varepsilon_r(d+\delta) + \varepsilon_{r1}h)}. \tag{25}$$

In this case, the sum of the charges on the outer electrodes is not independent of the membrane position.

C. Single ended

A single-ended transducer can be created by removing all of the structure to the right of the central strength membrane in Fig. 2 and replacing the resistor R on the right with a short circuit. In this case, $\sigma_{0b}=\sigma_{1b}=E_{gb}=E_{mb}=0$ and Eq. (7) becomes

$$-\varepsilon_0\varepsilon_r E_{ma}(z) = \sigma_{1a} - \rho_m z. \tag{26}$$

Substituting Eq. (5) in Eq. (26) yields

$$\sigma_{0a} + \sigma_{1a} - \sigma_m = 0. \tag{27}$$

In the quiescent state, where $e_{in}=0$ and $\delta=0$, Eq. (24) becomes

$$\sigma_{0a} = \frac{\varepsilon_{r1} h \sigma_m}{2(\varepsilon_r d + \varepsilon_{r1} h)}. \tag{28}$$

Referring to Fig. 3, it can be seen that as the gap width d is increased to infinity, σ_{0a} tends to zero, but σ_{1a} becomes asymptotically equal and opposite to the membrane charge $-\sigma_m$. Hence, the membrane, together with its conductive coating, forms a dipole which is its natural stable state. When $d=0$, the charge is shared equally between the two electrodes such that each holds half the value of the membrane charge.

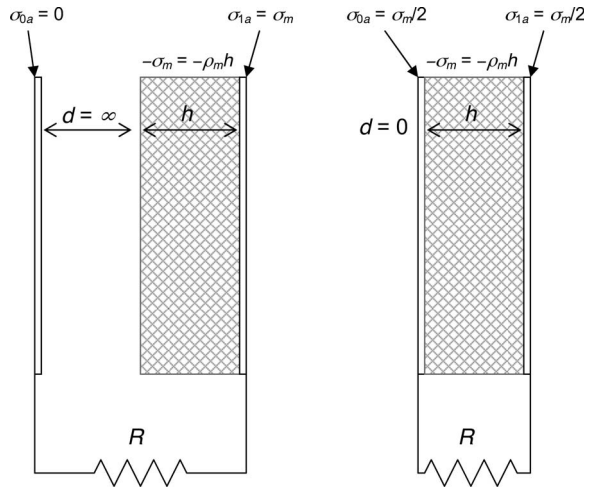


FIG. 3. Induced electrode charge at $d=0$ and $d=\infty$.

VIII. CALCULATION OF THE NET FORCE AS A FUNCTION OF VOLTAGE

A. Floating membrane

The net force is calculated from Eqs. (20)–(22). The charges σ_{1a} and σ_{1b} are the quiescent ones, in which state the membrane can be assumed to be grounded. They are taken from Eq. (9) as follows:

$$\sigma_{1a} + \sigma_{1b} = 2\sigma_m - \sigma_{0a} - \sigma_{0b}, \tag{29}$$

which, after inserting σ_{0a} and σ_{0b} from Eqs. (24) and (25), respectively, gives

$$\sigma_{1a} + \sigma_{1b} = 2\sigma_m - \frac{\varepsilon_{r1} h \sigma_m}{\varepsilon_r d + \varepsilon_{r1} h}, \tag{30}$$

which is assumed to be constant under dynamic conditions. After inserting Eqs. (21), (22), and (30) into Eq. (20), the net force F is then given by

$$F = -\frac{\varepsilon_r \varepsilon_{r1} h S \sigma_m}{2(\varepsilon_r d + \varepsilon_{r1} h)^2} e_{in} - \frac{\varepsilon_r \varepsilon_{r1} h^2 S \sigma_m^2}{2\varepsilon_0 (\varepsilon_r d + \varepsilon_{r1} h)^3} \delta. \tag{31}$$

There are two components to the force, which is perfectly linear. The first is due to the input voltage source and the second is due to a "negative stiffness" or the static attraction due to the membrane charge. Alternatively, Eq. (31) may be written as

$$F = \chi e_{in} + \kappa \delta, \tag{32}$$

where χ is the voltage-force conversion factor given by

$$\chi = \left(\frac{C_E}{C_M + C_G} \right) \frac{S\sigma_m}{2d} \approx \frac{\varepsilon_{r1} h S \sigma_m}{2\varepsilon_r d^2}, \quad d \gg \frac{\varepsilon_{r1} h}{\varepsilon_r}, \tag{33}$$

where C_E is the static capacitance between the outer electrodes when the membrane is blocked, which is given by

$$C_E = \frac{C_M C_G}{C_M + C_G}, \tag{34}$$

where C_M is the total capacitance of the two membranes given by

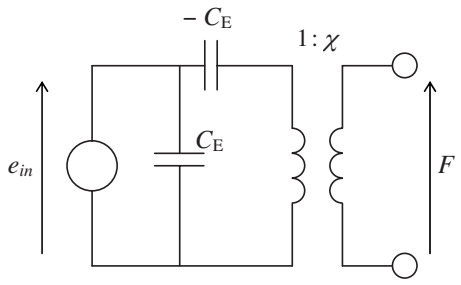


FIG. 4. Equivalent electrical circuit.

$$C_M = \frac{\varepsilon_0 \varepsilon_r S}{2h}, \tag{35}$$

and C_G is the total capacitance of the two gaps given by

$$C_G = \frac{\varepsilon_0 \varepsilon_{r1} S}{2d} \tag{36}$$

and κ is the negative stiffness given by

$$\kappa = \frac{C_E}{2(C_M + C_G)^2} \left(\frac{S\sigma_m}{d} \right)^2 \approx \frac{\varepsilon_{r1} h^2 S \sigma_m^2}{2\varepsilon_0 \varepsilon_r^2 d^3}, \quad d \gg \frac{\varepsilon_{r1} h}{\varepsilon_r}. \tag{37}$$

The equivalent electrical circuit is shown in Fig. 4, which can be used as part of a larger model including the dynamic impedance of the membrane and surrounding acoustic system. The force is plotted against displacement in Fig. 5 using Eq. (31) with $e_{in}=0$. It is interesting to compare this configuration with that of a nonelectret transducer[13] with an external polarizing voltage E_P, for which

$$\chi = \frac{E_P C_{ED}}{d} = \frac{\varepsilon_0 S E_P}{d^2}. \tag{38}$$

Hence, the external polarizing voltage for the equivalent nonelectret transducer is given by

$$E_P = \frac{S\sigma_m}{2(C_M + C_G)} \approx \frac{\sigma_m h}{2\varepsilon_0 \varepsilon_r}, \quad d \gg \frac{\varepsilon_{r1} h}{\varepsilon_r}. \tag{39}$$

Inserting σ_m from this equation into Eq. (31) indeed transforms it into Eq. (3.13) of Borwick[19] for the forces in a non-electret transducer.

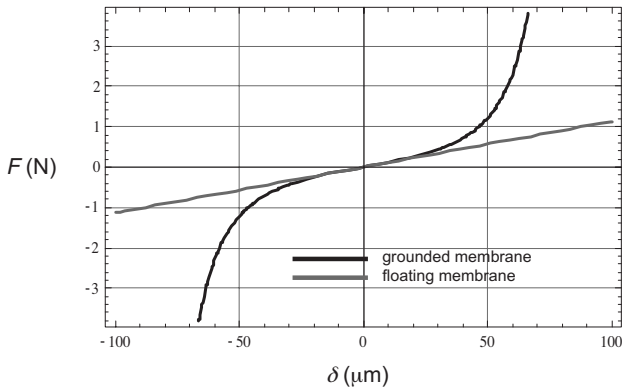


FIG. 5. Linearity of force vs displacement of push-pull transducer with input shorted where $S=\pi a^2$, where $a=15$ mm, $d=200$ μm, $h=12$ μm, $\sigma_m=14.1$ mC/m^2, $\varepsilon_0=8.85$ pF/m, $\varepsilon_r=12$, and $\varepsilon_{r1}=1$.

B. Grounded membrane

After inserting Eqs. (24) and (25) into Eq. (20), the net force F is then given by

$$F = \frac{\varepsilon_{r1} S}{8\varepsilon_0}\left\{\left(\frac{\varepsilon_0\varepsilon_r e_{in} - h\sigma_m}{\varepsilon_r(d+\delta)+\varepsilon_{r1}h}\right)^2 - \left(\frac{\varepsilon_0\varepsilon_r e_{in} + h\sigma_m}{\varepsilon_r(d-\delta)+\varepsilon_{r1}h}\right)^2\right\}. \tag{40}$$

Not surprisingly, this is the resultant force of two single-ended transducers mechanically coupled back to back. While the voltage to force conversion of this configuration is perfectly linear when $\delta=0$ (unlike a single-ended configuration), the force is nonlinear with displacement, as can be seen in Fig. 5, and it increases asymptotically as the membrane approaches each electrode where it is singular. The small-signal linearized expressions for the negative stiffness κ and voltage-force conversion factor χ can be obtained as follows:

$$\kappa = -\left.\frac{\partial F}{\partial \delta}\right|_{\delta=0, e_{in}=0} = \frac{\varepsilon_r\varepsilon_{r1}h^2 S\sigma_m^2}{2\varepsilon_0(\varepsilon_r d+\varepsilon_{r1}h)^3}, \tag{41}$$

$$\chi = \left.\frac{\partial F}{\partial e_{in}}\right|_{\delta=0, e_{in}=0} = \frac{\varepsilon_r\varepsilon_{r1}hS\sigma_m}{2(\varepsilon_r d+\varepsilon_{r1}h)^2}, \tag{42}$$

and these are the same as in Eq. (31) for the fully floating configuration.

C. Single ended

It can be shown that the net force F is then given by

$$F = \frac{\varepsilon_{r1} S}{8\varepsilon_0}\left(\frac{2\varepsilon_0\varepsilon_r e_{in} + h\sigma_m}{\varepsilon_r(d+\delta)+\varepsilon_{r1}h}\right)^2 = \frac{\varepsilon_{r1} S}{2(\varepsilon_r(d+\delta)+\varepsilon_{r1}h)^2}\left(\frac{h^2\sigma_m^2}{4\varepsilon_0} + \varepsilon_r h\sigma_m e_{in} + \varepsilon_0\varepsilon_r^2 e_{in}^2\right). \tag{43}$$

This equation is the same as that due to Sessler,[11] except that σ_m was multiplied by 2 in Sessler's model where a concentrated charge layer was assumed to exist at the interface between the membrane and the gap. It can be seen that there are three terms in the bracket. The first is the static force with zero input voltage. The other two are voltage dependent terms, the first of which is linear and the second, which depends upon the square of the input voltage, is nonlinear. It can also be seen that the force is nonlinear with displacement δ. Ignoring second order terms, the following linearized small-signal equation can be written as

$$F = \chi e_{in} + \kappa\delta, \tag{44}$$

where χ is the voltage to force conversion factor given by

$$\chi = \left(\frac{C_E}{C_M + C_G}\right)\frac{S\sigma_m}{2d} \approx \frac{\varepsilon_{r1}hS\sigma_m}{2\varepsilon_r d^2}, \quad d \gg \frac{\varepsilon_{r1}h}{\varepsilon_r} \tag{45}$$

and C_E is the static capacitance between the outer electrodes when the membrane is blocked, which is given by

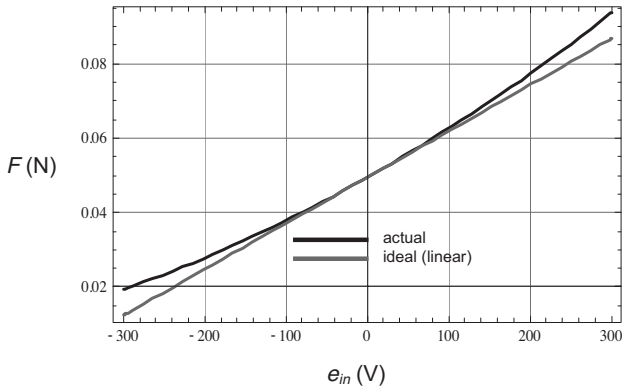


FIG. 6. Linearity of force vs voltage of single-ended transducer with membrane blocked where $S=\pi a^2$, where $a=15$ mm, $d=200$ μm, $h=12$ μm, $\sigma_m=14.1$ mC/m^2, $\varepsilon_0=8.85$ pF/m, $\varepsilon_r=12$, and $\varepsilon_{r1}=1$.

$$C_E = \frac{C_M C_G}{C_M + C_G}, \tag{46}$$

where C_M is the capacitance of the membrane given by

$$C_M = \frac{\varepsilon_0\varepsilon_r S}{h}, \tag{47}$$

and C_G is the capacitance of the gap given by

$$C_G = \frac{\varepsilon_0\varepsilon_{r1} S}{d}, \tag{48}$$

and κ is the negative stiffness. A small-signal linear approximation for κ is given by

$$\kappa = \frac{C_E}{4(C_M + C_G)^2}\left(\frac{S\sigma_m}{d}\right)^2 \approx \frac{\varepsilon_{r1}h^2 S\sigma_m^2}{4\varepsilon_0\varepsilon_r^2 d^3}, \quad d \gg \frac{\varepsilon_{r1}h}{\varepsilon_r}. \tag{49}$$

The equivalent electrical circuit is the same as that shown in Fig. 4 for a push-pull transducer. The force versus voltage characteristic of the blocked membrane is plotted in Fig. 6 using Eq. (43) with $\delta=0$. Also shown is the ideal linear case with the e_{in}^2 term omitted. The force versus displacement characteristic of the membrane with the electrodes shorted is plotted in Fig. 7 using Eq. (43) with $e_{in}=0$. Also shown is the ideal linear case using Eq. (49) added to the static force from Eq. (43) with $e_{in}=\delta=0$.

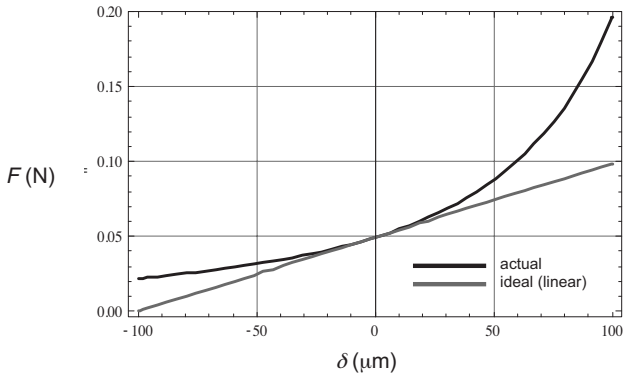


FIG. 7. Linearity of force vs displacement of single-ended transducer with input shorted and same parameters as in Fig. 6

D. Calculation of the net charge

In the case of a single-ended transducer, one can calculate the net stored charge by applying a reverse voltage such that the membrane displacement is reduced and then increasing it further until the displacement is about to increase again. Let the turning point occur at voltage e_0. Then differentiating Eq. (43) with respect to e_{in} gives

$$\left.\frac{dF}{de_{in}}\right|_{e_{in}=e_0} = \frac{\varepsilon_{r1}S(\varepsilon_r h\sigma_m + 2\varepsilon_0\varepsilon_r^2 e_{in})}{2(\varepsilon_r(d+\delta)+\varepsilon_{r1}h)^2} = 0, \tag{50}$$

so that the surface charge density is given by

$$\sigma_m = -\frac{2\varepsilon_0\varepsilon_r e_0}{h}. \tag{51}$$

By inserting this expression for σ_m in Eq. (43), it can be shown that $F=0$ when $e_{in}=e_0$, so there should be zero displacement at this point.

IX. STABILITY OF A CIRCULAR MEMBRANE

For a circular membrane, the following static wave equation can be written, ignoring inertia and external loading,

$$T\left(\frac{\partial^2}{\partial w^2} + \frac{1}{w}\frac{\partial}{\partial w}\right)\eta(w) - \frac{\kappa}{\pi a^2}\eta(w) = 0, \quad \eta(a) = 0, \tag{52}$$

where η is the membrane deflection, w is the radial ordinate, T is the tension, and a is the radius. The solution is given by

$$\eta(w) = \begin{cases} 0, & a\sqrt{\kappa/(\pi a^2 T)} \neq \alpha_n \\ CJ_0(\alpha_n w/a), & a\sqrt{\kappa/(\pi a^2 T)} = \alpha_n, \end{cases} \tag{53}$$

where α_n is the nth zero of J_0, C is an arbitrary constant, and the square-root term is the static wave number. Hence, for stability

$$T > \frac{\kappa}{\pi\alpha_1^2}, \tag{54}$$

where $\alpha_1 = 2.4048$. Using Eq. (37) for κ, where the area is given by $S=\pi a^2$, and assuming $d \gg \varepsilon_{r1}h/\varepsilon_r$ give

$$T > \frac{a^2\varepsilon_{r1}h^2\sigma_m^2}{2\alpha_1^2\varepsilon_0\varepsilon_r^2 d^3}, \quad d \gg \frac{\varepsilon_{r1}h}{\varepsilon_r}, \tag{55}$$

for a push-pull transducer. In the case of a single-ended transducer, this value of tension is halved. Inserting σ_m from Eq. (39) into Eq. (55) gives

$$T > \frac{2a^2\varepsilon_0\varepsilon_{r1}E_P^2}{\alpha_1^2 d^3}, \tag{56}$$

which is Streng's equation[12] for the stability of a nonelectret push-pull electrostatic transducer. It can be seen that as the charge density is increased, the tension also has to be increased in order to maintain stability. However, the fundamental resonant frequency of the membrane has to be set low enough to give the desired bandwidth and this is determined by the system stiffness. The system stiffness is that due to the membrane (which is proportional to its tension) less the negative stiffness due to the electrostatic force. As the charge density is increased, the difference between these two forces becomes an ever smaller proportion of the membrane stiffness. Hence, a small relaxation in its tension, which can easily occur due to age or environmental conditions, can lead to the membrane becoming attached to either of the electrodes. Some solutions have been proposed,[20] which involve increasing the stiffness of the membrane by adding "springs" between it and the electrodes. Unfortunately, these add nonlinearity to an otherwise linear transducer. Inevitably, they restrict the maximum displacement and hence also the maximum sound pressure. On the other hand, they can enable new forms such as flexible loudspeakers.[21]

X. DISCUSSION OF THE RESULTS

In all of the configurations considered, there is a near field effect whereby for very small electrode separations (i.e., $d < h\varepsilon_{r1}/\varepsilon_r$), the force is relatively independent of the separation. However, such a small separation is not such a practical proposition for a loudspeaker since the membrane would not be able to move far enough to produce a reasonable sound pressure. At larger distances, these expressions reduce to simpler far-field expressions in which the driving force is proportional to the inverse square of the separation and the negative stiffness is proportional to the inverse cube of the separation.

The floating push-pull electret transducer is utterly linear with both input voltage and displacement, as can be seen from Fig. 5, because the charges are kept constant by virtue of the floating membrane. However, it should be noted that under static conditions, as well as at frequencies below $f = 1/(2\pi RC_E)$, this configuration behaves as though the membrane is grounded due to charge leakage. Even if there are no physical resistors R connected to the conductive coatings, as shown in Fig. 2, those resistors will be replaced by charge leakage paths so that the static deflection is still nonlinear, as also shown in Fig. 5.

Interestingly, although the driving force of the grounded membrane configuration is nonlinear with displacement, it is linear with input voltage when the membrane is locked in its central resting position. This is due to the fact that the charges induced in the membrane's conductive coatings vary with the position of the membrane (which is also true for the single-ended configuration). Not surprisingly, it can be seen from Figs. 6 and 7 that the single-ended configuration yields a driving force that is neither linear with input voltage nor displacement.

XI. CONCLUSIONS

The force-versus-voltage and force-versus-displacement sensitivities of the electret transducer have been derived, assuming the charge to be evenly distributed throughout the membrane dielectric. Using the same derivation method, it can be shown that this is equivalent to having a concentrated charge layer in the middle. Thus, any symmetrical charge distribution can be represented by this model. Furthermore, this concentrated charge layer can be regarded as one plate of a capacitor, with the conductive coating forming the other. Hence, the external polarizing voltage for an equivalent non-

electret transducer is the same as that across this notional capacitor and is related to the charge by Eq. (39), which is essentially the total stored charge divided by the capacitance of the half membrane. The electromechanical force conversion factor has also been shown on an equivalent circuit, together with the interelectrode capacitance and negative capacitance, as a basis for simulation.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Dr. Eino Jakku and Dr. Dar-Ming Chiang for their invaluable discussions during the preparation of the manuscript.

[1]G. M. Sessler and J. E. West, "Self-biased condenser microphone with high capactiance," J. Acoust. Soc. Am. **34**, 1787–1788 (1962).
[2]G. M. Sessler, "Electrostatic microphones with electret foils," J. Acoust. Soc. Am. **35**, 1354–1357 (1963).
[3]G. M. Sessler and J. E. West, "Studies of electret charges produced on polymer films by electron bombardment," Polym. Lett. **7**, 367 (1969).
[4]G. M. Sessler and J. E. West, "Electret transducers: A review," J. Acoust. Soc. Am. **53**, 1589–1600 (1973).
[5]Z. Xia, A. Wedel, and R. Danz, "Charge storage and its dynamics in porous polytetrafluoroethylene (PTFE) film electrets," IEEE Trans. Dielectr. Electr. Insul. **10**, 102–108 (2003).
[6]D.-M. Chiang, W.-L. Liu, J.-L. Chen, and R. Susuki, "PALS and SPM/EFM investigation of charged nanoporous electret films," Chem. Phys. Lett. **412**, 50–54 (2005).
[7]F. Paschen, "Ueber die zum Funkenübergang in Luft, Wasserstoff und Kohlensäure bei verschiedenen Drucken erforderliche Potentialdifferenz (On the potential difference required to cause spark-over in air, hydrogen and carbon dioxide under different pressures)," Ann. Phys. **273**(5), 69–96 (1889).
[8]J. R. Wright, "The virtual loudspeaker cabinet," J. Audio Eng. Soc. **51**, 244–247 (2003).
[9]T. J. Mellow, O. Umnova, K. Drossos, K. R. Holland, A. J. Flewitt, and L. M. Kärkkäinen, "On the adsorption-desorption relaxation time of carbon in very narrow ducts," J. Acoust. Soc. Am. **123**, 3141 (2008).
[10]F. V. Hunt, *Electroacoustics* (Acoustical Society of America, New York, 1982).
[11]G. M. Sessler, "Electric fields and forces due to charged dielectrics," J. Appl. Phys. **43**, 405–408 (1972).
[12]J. H. Streng, "Sound radiation from a circular stretched membrane in free space," J. Audio Eng. Soc. **37**, 107–118 (1989), Eqs. (19) and (21).
[13]T. J. Mellow and L. M. Kärkkäinen, "On the sound field of a circular membrane in free space and an infinite baffle," J. Acoust. Soc. Am. **120**, 2460–2477 (2006).
[14]T. J. Mellow, "On the sound field of a resilient disk in free space," J. Acoust. Soc. Am. **123**, 1880–1891 (2008).
[15]T. J. Mellow and L. M. Kärkkäinen, "Comparison of spheroidal and eigenfunction-expansion trial functions for a membrane in an infinite baffle," J. Acoust. Soc. Am. **123**, 2598–1602 (2008).
[16]J. E. Warren, "Capacitance microphone static membrane deflections: Comments and further results," J. Acoust. Soc. Am. **58**, 733–740 (1975).
[17]N. Atoji and T. Aoi, "Electrostatic acoustic transducer," U.S. Patent No. 3, 833,770 (3 September 1974).
[18]H. Koskinen and A. Viljanen, "Electrodynamiikka (Electrodynamics)," The course notes on electrodynamics from University of Helsinki, 2006, http://theory.physics.helsinki.fi/~ed/luento/ED2006.pdf. Last viewed online on 15 May 2008.
[19]J. Borwick, *Loudspeaker and Headphone Handbook*, 3rd ed. (Focal, Oxford, 2001), p. 112, Eq. (3.13).
[20]L. A. Whitehead, "High pressure low impedance electrostatic transducer," U.S. Patent No. 5,450,498 (12 September 1995). A brief review by D. W. Martin appears in J. Acoust. Soc. Am. **100**(1), 21 (1996).
[21]D.-M. Chiang, and J.-L. Chen, "A novel flexible loudspeaker driven by an electret diaphragm," Audio Engineering Society Convention Papers, 121st Convention, San Francisco, 5-8 October 2006, available from Audio Engineering Society Inc. (unpublished).

A note on determination of the diffuse-field sensitivity of microphones using the reciprocity technique[a)]

Salvador Barrera-Figueroa[b)] and Knud Rasmussen[c)]
Danish Primary Laboratory of Acoustics, Danish Fundamental Metrology, Matematiktorvet B307, DK-2800 Lyngby, Denmark

Finn Jacobsen[d)]
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads B352, DK-2800 Lyngby, Denmark

(Received 12 December 2007; revised 29 May 2008; accepted 30 May 2008)

The diffuse-field response of a microphone is usually obtained by adding a random-incidence correction to the pressure response of the microphone. However, the random-incidence correction is determined from a relative measurement, and its accuracy depends not only on the relative response at all angles of incidence but also on the accuracy of the frequency response at normal incidence. By contrast, this paper is concerned with determining the absolute diffuse-field response of a microphone using the reciprocity technique. To examine this possibility, a reciprocity calibration setup is used for measuring the electrical transfer impedance between a pair of microphones placed in a miniature (2 m^3) reverberation room. The transfer function between the microphones is measured using fast Fourier transform analysis and pseudorandom noise. The calculation of the diffuse-field sensitivity involves (a) separation of the reverberant response from the total response, (b) determination of the reverberation time, and (c) averaging over space and frequency. The resulting diffuse-field correction is compared with an estimate of the random-incidence correction determined in an anechoic room and with a numerical prediction.
 [DOI: 10.1121/1.2950089]

PACS number(s): 43.38.Kb, 43.58.Vb [AJZ] Pages: 1505–1512

I. INTRODUCTION

The sensitivity of a microphone can be defined in terms of the sound field in which it is immersed. Uniform pressure conditions imply that the membrane of the microphone is exposed to a uniform sound pressure, for example, in a small coupler. Under free-field conditions, a plane wave impinges on the body of the microphone and the response of the microphone depends on the scattering caused by the rigid surfaces of the microphone and its radiation impedance. Finally, under ideal diffuse-field conditions, an infinite number of incoherent plane waves coming from random directions and having random phases and amplitudes impinges on the microphone; thus the diffuse-field sensitivity is the response of the microphone in a "perfectly diffuse sound field," an unrealizable, idealized concept that corresponds to an average over an infinite ensemble of realizations of sound fields composed of infinitely many plane waves with random phases arriving from all directions.[1] Although such a sound field can be approximated in reverberation rooms, another quantity, the random-incidence sensitivity, is used conventionally as an equivalent to the diffuse-field sensitivity. The random-incidence sensitivity is the incoherent combination of the responses of the microphone at all possible directions. In practice this quantity is determined by averaging the response of the microphone to waves coming sequentially from all directions.[2]

Today practically all microphones are calibrated using pressure-field methods, although more than 90% of all microphones are used under diffuse-field or free-field conditions. The diffuse-field sensitivity of a microphone is in practice determined from its measured pressure sensitivity using empirical corrections that depend on the type and geometry of the microphone.[3] The disadvantage of this method is that it gives rise to long delays when new transducers are introduced because it takes time for such corrections to be recognized internationally. Another problem is that some transducers, e.g., microelectromechanical system (MEMS) microphones, simply cannot be pressure calibrated.

The reciprocity technique for microphone calibration under pressure-field conditions in closed couplers was introduced in the 1940s,[4–7] and has since then been developed to a level of very high accuracy. The reciprocity technique yields the absolute sensitivity, and thus it is used in primary calibration.[8] The free-field reciprocity technique was also introduced in the 1940s,[6,7] but it took much longer time to develop this method to a high level of accuracy,[9] partly because of the extreme difference in signal levels (if the transmitter microphone is driven with 10 V the response from the receiver microphone will typically be less than 1 μV) and partly because anechoic rooms in practice are less than perfect combined with the fact that reflections have a serious

[a)]Portions of this work were presented in "An investigation of microphone calibration in a diffuse sound field," Proceedings of Inter-Noise 2007, Istanbul, Turkey, August 2007.
[b)]Electronic mail: sbf@dfm.dtu.dk
[c)]Electronic mail: kra@dfm.dtu.dk
[d)]Electronic mail: fja@elektro.dtu.dk

influence. More recently, time-selective techniques based on advanced signal processing have been developed.[10] Application of the time-selective techniques implies the possibility of doing free-field calibrations under nonanechoic conditions. Only a little has been published on diffuse-field calibration of microphones.[11–13] The only fundamental studies were published about 40 years ago,[11,12] and no standard for diffuse-field calibration of microphones has ever been developed, undoubtedly because diffuse-field reciprocity calibration of microphones is even more difficult than free-field reciprocity calibration.

The random-incidence sensitivity is conventionally determined from a pressure calibration combined with a random-incidence correction that is based on the free-field correction and an additional correction for nonaxial sound incidence, where the latter is obtained from relative measurements of the frequency response of the microphone in a free field when a succession of sound waves impinges on the microphone from all directions.[2,14,15] Alternatively, it could be determined from a free-field calibration combined with the correction for nonaxial incidence. However, direct, absolute diffuse-field calibration would shorten the measurement chain. The purpose of this investigation is to establish a setup for reciprocity calibration of condenser microphones under diffuse-field conditions, to examine the sound field statistics, and to demonstrate that diffuse-field reciprocity calibration is possible with high precision. The investigated issues are (i) measurement of the transfer function between microphones in a reverberation room, (ii) separation of the direct and the reverberant responses using a time-selective technique, (iii) calculation of the diffuse-field sensitivity of laboratory standard (LS) microphones, (iv) averaging of the resulting sensitivity over frequency and space, and (v) validation of the diffuse-field response using a random-incidence estimate.

II. THE DIFFUSE-FIELD SENSITIVITY

Formally, the diffuse-field sensitivity is defined as follows: "for a sinusoidal signal of a given frequency in a diffuse sound field and for given environmental conditions, the quotient of the open-circuit voltage of the microphone by the sound pressure that would exist at the position of the acoustic center of the microphone in the absence of the microphone."[16] However, a sinusoidal signal inevitably implies an interference field with large spatial variations of the sound pressure level, in obvious disagreement with the concept of a perfectly diffuse sound field.[1,17–19] Diestel[11] and Nakajima[12] tacitly used a far more reasonable definition that, in fact, corresponds to the ensemble average of realizations of such interference fields, in other words, to the random-incidence sensitivity.

The electrical transfer impedance between two microphones in a diffuse field is[11]

$$Z_{12} = \frac{u_2}{i_1} = M_{d,1} \cdot M_{d,2} \frac{1}{J_{d,12}}, \tag{1}$$

where u_2 is the open-circuit voltage on the terminals of the receiver microphone, i_1 is the current through the terminals of the transmitter microphone, $M_{d,1}$ and $M_{d,2}$ are the diffuse-field sensitivities of microphones 1 and 2, respectively, and $J_{d,12}$ is the reciprocity factor. The reciprocity factor, which follows from the relation between sound power emitted by the transmitter microphone and the resulting mean-square pressure in the reverberation room,[11] is

$$J_d = \left(\frac{6}{\pi \log e}\right)^{1/2} \left(\frac{1}{\rho_0 f}\right) \left(\frac{V}{cT_{60}}\right)^{1/2}, \tag{2}$$

where V is the volume of the reverberation room, c is the speed of sound, T_{60} is the Sabine reverberation time, f is the frequency, and ρ_0 is the density of air. If a third microphone is coupled sequentially with microphones 1 and 2, then the absolute diffuse-field sensitivities of the three microphones can be determined by solving three equations that contain the three products of sensitivity. For instance, the diffuse-field sensitivity of microphone 1 can be obtained from

$$M_{d,1}^2 = \frac{Z_{12} Z_{13}}{Z_{23}} \frac{J_{d,12} J_{d,13}}{J_{d,23}}. \tag{3}$$

Similar equations can be derived for microphones 2 and 3. A further simplification can be made if the reciprocity factor can be considered to be the same in the three measurements. Note, however, that the air absorption and thus the reverberation time, the speed of sound, and the density of air might change because of changes in the environmental conditions. Finally, the diffuse-field correction C_d can be defined as the logarithmic ratio of the diffuse-field sensitivity to the pressure sensitivity,

$$C_d = 10 \log \left| \frac{M_d}{M_p} \right|^2, \tag{4}$$

where M_p is the pressure sensitivity determined by reciprocity calibration (or from the electrostatic actuator response and the low-frequency sensitivity from a piston-phone calibration). This quantity will be compared with the random-incidence correction.

It should be added that the sound power emitted by a pure tone source in a reverberation room varies significantly with the position of the source and with the frequency at low frequencies,[20,21] from which it follows that Eq. (2) is only valid on average: source position averaging is needed.

III. REALIZATION OF A DIFFUSE SOUND FIELD

It is a normal practice in pressure and free-field reciprocity calibration that the signals used are pure sinusoidal tones. This is particularly appropriate in free-field calibration because the signal-to-noise ratio is best using such a signal. In pressure calibration, signal levels are not the most important issue because the small cavities used normally give good signal-to-noise ratios. In principle, it is also possible to use pure tones in a reverberation room. The sound field in the room will be the product of a large number of interferences between reflected waves. This sound field has well defined statistical characteristics that make it possible to estimate, e.g., the spatial variation of the sound pressure in such a sound field. Above the Schroeder frequency,[22] the mean-square pressure in a reverberant sound field generated by a pure tone, $\overline{p^2}$, may be regarded as a stochastic process which

is a random function of the frequency[22,23] and a random function of the position.[18,19] Both functions have an exponential distribution. The normalized variance of the exponential distribution, that is, the ratio of the variance to the square of the mean, $\sigma^2(\overline{p^2})/E^2(\overline{p^2})$, equals 1. This implies that the variations of the sound pressure with position and with frequency are very large.

The electrical transfer impedance is directly proportional to the sound pressure impinging on the receiver microphone. Thus, the resulting sensitivity will have similar statistical properties as the sound field generated by a pure tone source. On the other hand, if the reverberant response is to be separated from the total response, the length of the impulse response that corresponds to the measured transfer impedance as a function of the frequency must include the whole reverberant response. This requires a very fine sampling in the frequency domain that would make the application of sequential pure tone measurements prohibitively time consuming. A suitable solution is to use pseudorandom noise instead. Because of the fact that pseudorandom noise is a multitone signal, the resulting sound field will, at any frequency, have the same characteristics as the sound field generated by a pure tone. Therefore the use of pseudorandom noise reduces the measurement time dramatically. However, any signal processing technique capable of determining a sufficiently long impulse response will do; one might, for example, use frequency sweeps.[24]

A. Spatial averaging

In general, the variance of the average of a random variable, in this case the mean-square pressure (or the mean-square sensitivity) averaged over n positions in the room, $\mathbf{r}_i$, is given by

$$\sigma^2\left\{\frac{1}{n}\sum_{i=1}^{n}\overline{\mathbf{p}^2}(\mathbf{r}_i)\right\} = \frac{1}{n}\sigma^2(\overline{\mathbf{p}^2}) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq 1}^{n}\mathrm{cov}\{\overline{\mathbf{p}^2}(\mathbf{r}_i),\overline{\mathbf{p}^2}(\mathbf{r}_j)\}. \quad (5)$$

If the positions represent independent samples of the stochastic process (which means that they should be spaced at least by half a wavelength[25,26]) the second term on the right side of Eq. (5) vanishes, and the resulting variance is inversely proportional to the number of positions. It follows that the standard deviation will decrease with the inverse of the square root of n. Considering that the normalized variance of an exponential distribution is equal to 1, it follows that it is necessary to average over a very large number of independent positions if a small variance of the spatial average is required.

B. Frequency averaging

In the same way as with the spatial averaging, the variance of the mean-square pressure over a number of adjacent frequencies will decrease with the inverse of the number of independent samples. The frequency separation between independent samples depends on the losses of the room. The separation should satisfy[23]

$$\Delta f > \frac{6}{T_{60}}. \quad (6)$$

The mean-square pressure is also exponentially distributed when regarded as a function of the frequency. Therefore it is necessary to average over a large number of independent frequencies in order to obtain a small variance of the frequency average. With averaging over a wide band of width B the variance is reduced by a factor of $BT_{60}\log e/3 \approx BT_{60}/6.9$.[19,27]

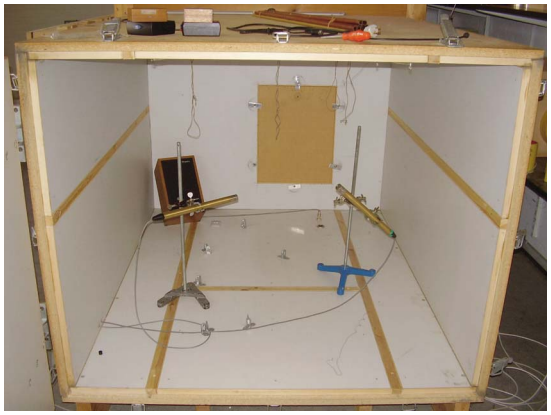


FIG. 1. (Color online) The reverberation chamber with the measurement rig. The loudspeaker is used for measuring the reverberation time.

IV. EXPERIMENTAL SETUP

A. Realizing the diffuse field

If the measurements are carried out in a large reverberation room, the sound pressure level generated by the transmitter microphone will be low and it will be necessary to use very long microphone cables, which may have an adverse effect on the signal-to-noise ratio. Therefore, the measurements were carried out in a miniature reverberation room [a scale model (1:5) of the reverberation rooms at DTU]. The reverberation chamber has a volume of approximately 2 m^3; see Fig. 1. The reverberation time of the room was measured in one-twelfth octave bands with a Brüel & Kjær (B&K) PULSE analyzer using the conventional interrupted noise method and averaging over repeated decays (five decays at five positions). The reverberation time and the associated spatial standard deviation are presented in Fig. 2. The Schroeder frequency of the room was estimated to be about 1 kHz. This is perfectly adequate for microphone calibration because the difference between the diffuse-field sensitivity and the pressure sensitivity of a microphone is very small below 1 kHz.

B. Measurement system

The measurement setup was composed of a B&K reciprocity apparatus of type 5998, a microphone amplifier B&K type NEXUS model 2690, and a B&K PULSE multichannel analyzer. The receiver microphone was connected to a pre-

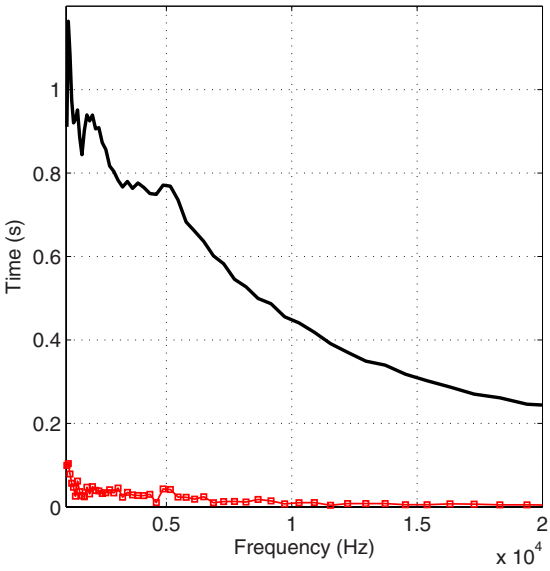


FIG. 2. (Color online) The reverberation time measured in the reverberation chamber. The thick solid line is the reverberation time and the thin solid line is the spatial standard deviation.

amplifier with a 20 dB built-in amplifier. The influence of the load of the preamplifier was eliminated using the conventional insert voltage technique.[8] The preamplifiers were attached to cylindrical rods of the same diameter as the microphones and sufficiently long to approximate semi-infinite cylinders in the frequency range of concern. Figure 3 shows a block diagram of the measurement setup. The ratio of receiver-transmitter output voltages was measured with fast Fourier transform (FFT) analysis using synchronized pseudorandom noise. Determination of the impulse response requires a fine frequency sampling of the frequency response. Thus, it was necessary to measure the frequency response in frequency steps of $\Delta f = 0.5$ Hz. In order to cover the entire frequency range with such fine resolution sequential "zoom FFT" measurements were carried out. Each zoom had a bandwidth of 3.2 kHz and 6400 frequency lines.

C. Separation of the reverberant response

The separation of the direct and reverberant response can be carried out using a time-selective procedure very similar to the one described in Ref. 10. The procedure applied in this case was as follows: (a) The transfer function between the signals of two microphones was measured; (b) at low and high frequencies this frequency response was taken smoothly to zero by applying a bandpass filter; (c) an inverse discrete Fourier transform was calculated; (d) a time-selective window was applied to the resulting impulse response in order to separate the direct and the reverberant response; and (e) a discrete Fourier transform was applied to the reverberant part of the impulse response. The transfer function in (a) was measured using the procedure and parameters described in the previous subsection. The bandpass filter used in (b) had roll-off frequencies of 2 and 23 kHz. The time-selective window used in (d) was a Tukey window with a length of 1.99 s, and the percentage of the smoothing portion at the ends of the window was of 0.1% of the length of the window. The Tukey window was selected because it does not modify the impulse response beyond the smoothing at

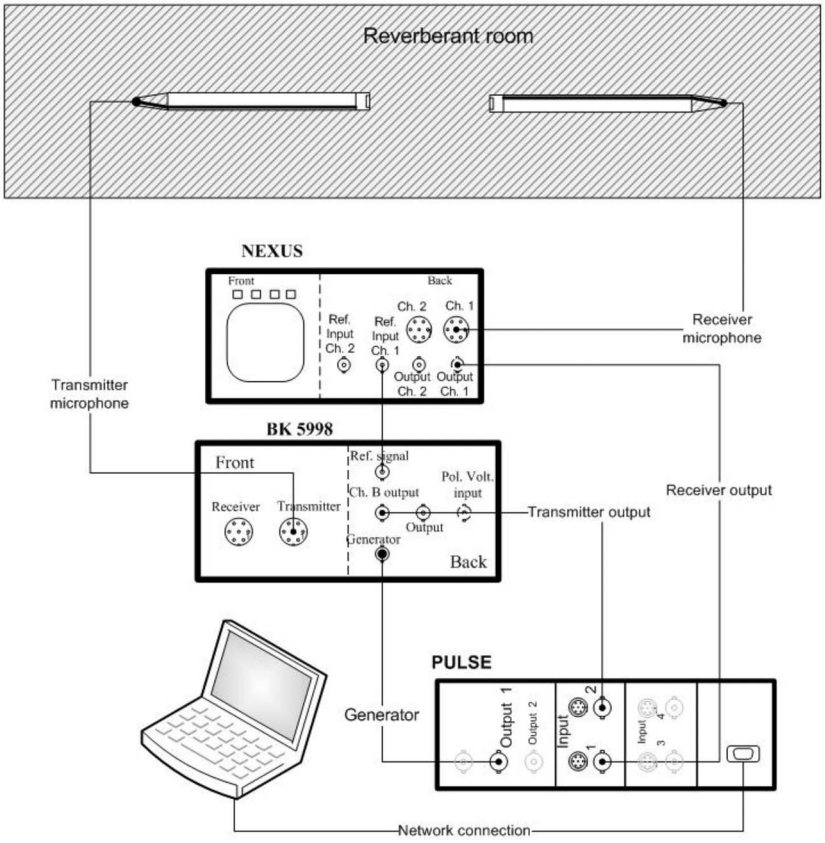


FIG. 3. Block diagram of the measurement system.

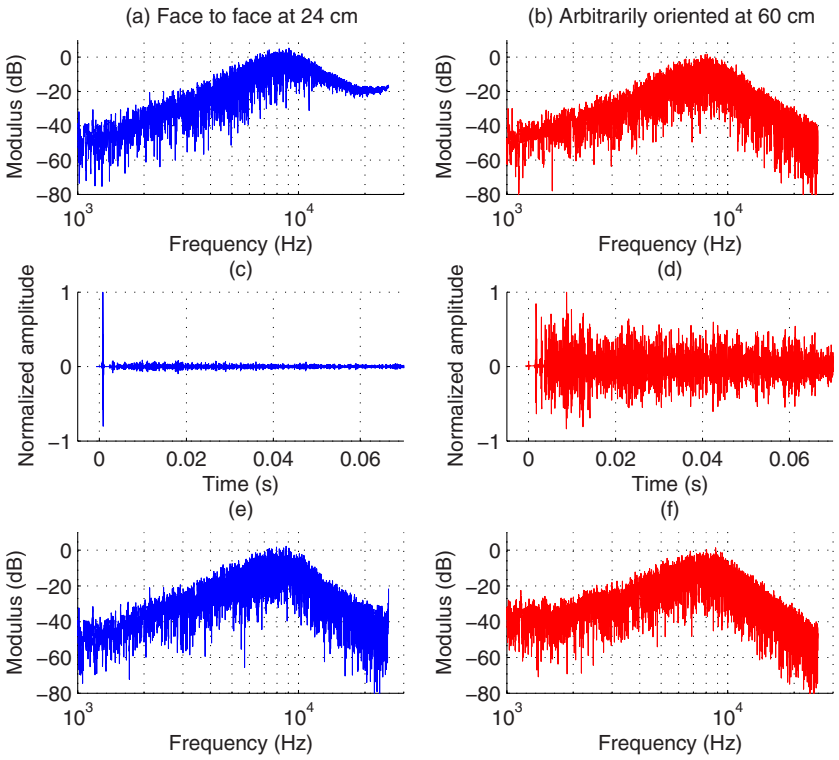


FIG. 4. (Color online) Frequency and impulse responses measured with different configurations in the reverberation chamber. Graphs in the left column display data obtained from measurements of the transfer function with the microphones face to face at a distance of 24 cm from each other; graphs in the right column display data obtained from measurements of the transfer function with the microphones oriented at arbitrary directions and at approximately 60 cm from each other. Graphs (a) and (b) show the measured ratios of voltages, graphs (c) and (d) are the impulse responses, and graphs (e) and (f) are the reverberant ratios of voltages.

the ends of the window; the window starts at 0.002 s. In this way, the reverberant impulse response was separated from the direct response.

V. EXPERIMENTAL RESULTS

Figure 4 shows the transfer impedance between two 1 in. laboratory standard (LS1) microphones measured at two different sets of positions in the reverberant room: in one case the microphones were face to face at a distance of 24 cm from each other, and in the other case the microphones had arbitrary orientations and were placed at a distance of approximately 60 cm from each other.

The fluctuations of the frequency response shown in Fig. 4(a) start to decrease significantly above 10 kHz when the microphones face each other at a short distance. This is a consequence of the high ratio of the direct field to the reverberant field in this frequency range, and this can be expected because LS1 microphones are highly directional. By contrast the fluctuations of the transfer impedance between the microphones oriented at arbitrary directions do not decrease at high frequencies. The impulse responses shown in Figs. 4(c) and 4(d) confirm this interpretation; in Fig. 4(c) the direct impulse response is much larger than the reverberant response whereas in Fig. 4(d) the two responses are of similar amplitude. Finally, the separated reverberant responses shown in Figs. 4(e) and 4(f) were determined using the procedure described above. As can be seen, the reverberant responses are very similar for the two cases under consideration. This indicates that it is possible to measure at any position in the reverberation chamber and to separate the reverberant response from the direct sound or early reflections, leaving only the reverberant part to be used for further analysis.

The next step is to calculate the diffuse-field sensitivity from the reverberant response. Figure 5 shows the diffuse-field sensitivity of an LS1 microphone. The calculated sensitivity, which complies with the definition in Ref. 16, fluctuates dramatically with the frequency. These fluctuations are a consequence of the realization of the sound field and not a property of the microphone itself. Thus, in order to reduce the fluctuations to an acceptable level the sensitivity must be averaged over different positions in the chamber and over frequency. It was also mentioned in Sec. II that the reciprocity factor is only valid on average. According to Eq. (5) one would have to average over about 1850 independent positions in the reverberation chamber in order to reduce the

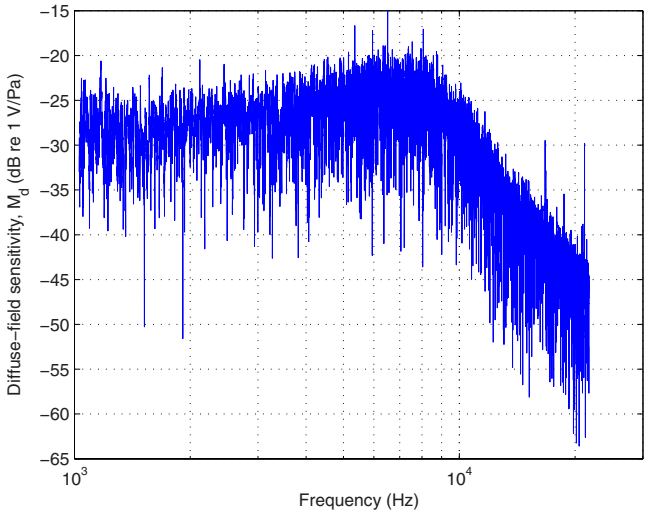


FIG. 5. (Color online) Diffuse-field sensitivity of an LS1 microphone calculated from the reverberant transfer function for a single microphone configuration using the reciprocity technique.

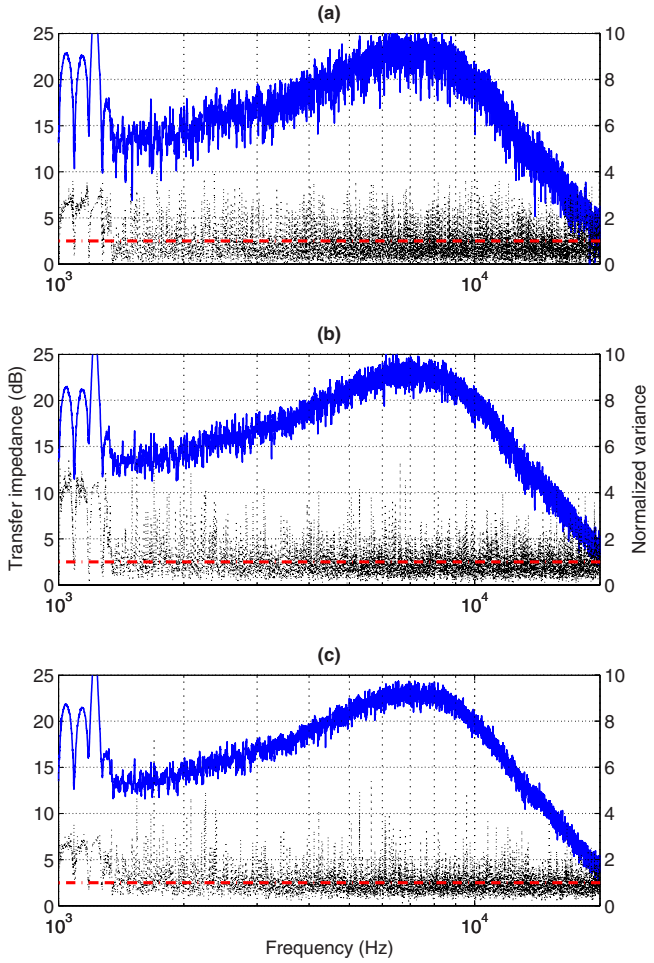


FIG. 6. (Color online) Spatial average and normalized variance of the transfer impedance between two LS1 microphones determined from measurements in the reverberation chamber at (a) 4 different positions, (b) 9 different positions, and (c) 16 different positions. The thick solid line is the modulus of the transfer impedance (in decibels), the thin dotted line is the normalized variance, and the thick dash-dotted line is the predicted normalized variance.

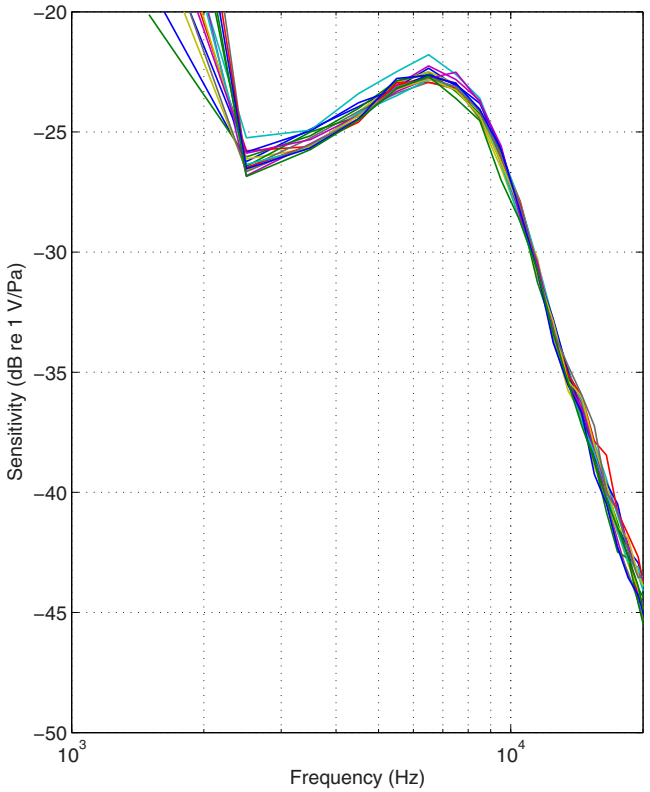


FIG. 7. (Color online) Frequency averaged diffuse-field sensitivity of an LS1 microphone determined from measurements at 16 different configurations in the scale model of a reverberation room.

standard deviation of the mean-square sensitivity to 0.1 dB. This is prohibitive, not only from a practical point of view but also because there simply are not so many independent positions in the chamber. This leads to the conclusion that frequency averaging is also needed.

Before proceeding, it should be decided whether the averaging should be made before or after the calculation of the sensitivity. Averaging after the sensitivity is calculated (as in Fig. 5) requires that the fine structure of the transfer impedance is exactly the same for the three microphone combinations; thus the ratio $R=Z_{ij}Z_{ik}/Z_{jk}$ should be similar to the three transfer impedances. However, this may not always be the case; differences between the transfer impedances' "fine structure" may occur because of temperature changes and small random positioning errors that cannot easily be avoided when manual positioning is used. A more robust approach is to do the averaging over frequency of each squared transfer impedance *before* the ratio R is calculated. This is the approach adopted here.

Figure 6 shows the spatial average and the normalized variance of the squared modulus of the transfer impedance between microphones 1 and 2, determined from measurements at 4, 9, and 16 different positions in the reverberation chamber. Although the fluctuations decrease visibly when the number of averages increases, it is still clear that a much larger number of measurements at independent positions in the chamber would be required if a smoother realization of the final diffuse-field sensitivity is required. The value of the normalized variance oscillates around 1. This is consistent with the assumption that the squared transfer impedance is proportional to the mean-square pressure and therefore has an exponential distribution.

Figure 7 shows the results of averaging the diffuse-field sensitivity of an LS1 microphone measured at 16 random pairs of positions in the reverberation room over frequency. For each configuration the three squared transfer impedances are averaged over frequency bands with a width of 1 kHz, as described above. It can be seen that the fluctuations disappear almost completely. However, there are still differences between the sensitivities obtained from different microphone positions in the reverberation room. In addition to the uncertainty due to the variation with position of the sound power output of the transmitter microphone,[20,21] these differences might be caused, e.g., by changes of the environmental conditions during the measurements. However, the most important factor is undoubtedly the variability of the sound field itself. The spatial averaging will reduce the standard deviation by a factor of $\sqrt{n}$ [see Eq. (5)], and the frequency averaging will reduce it in inverse proportion to $\sqrt{BT_{60}}$, as mentioned above.

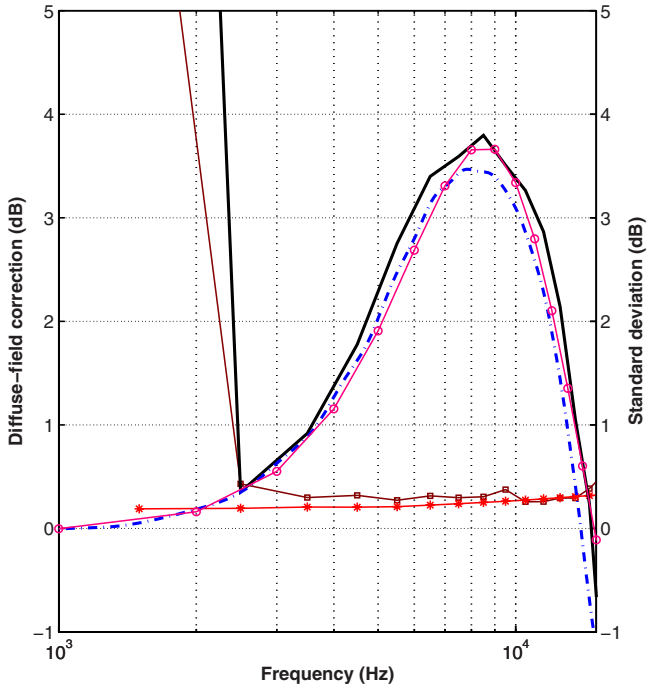


FIG. 8. (Color online) Diffuse-field correction of LS1 microphones. The thick solid line is the average diffuse-field correction, the dash-dotted line is the random-incidence correction from Ref. 15, the line with circular markers is the random-incidence correction calculated using the boundary element method, the line with square markers is the measured spatial standard deviation of the diffuse-field correction, and the line with star markers is the predicted spatial standard deviation of the diffuse-field sensitivity.

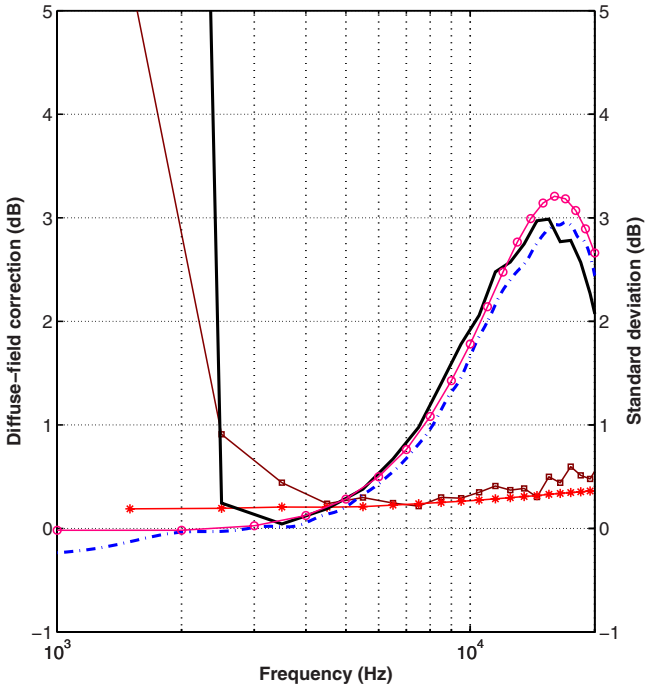


FIG. 9. (Color online) Diffuse-field correction of LS2 microphones. The thick solid line is the average diffuse-field correction, the dash-dotted line is the random-incidence correction from Ref. 15, the line with circular markers is the random-incidence correction calculated using the boundary element method, the line with square markers is the measured spatial standard deviation of the diffuse-field correction, and the line with star markers is the predicted spatial standard deviation of the diffuse-field sensitivity.

Finally, Figs. 8 and 9 show the average diffuse-field correction of three LS1 microphones and three LS2 (0.5 in.) microphones, respectively. The correction has been determined from the diffuse-field sensitivity averaged both over frequency and 16 sets of microphone positions in the reverberation chamber. The estimate of the diffuse-field correction is compared with estimates of the random-incidence correction from Ref. 15 and a numerical calculation determined using the boundary element method (BEM). The velocity distributions of the membranes of LS1 and LS2 microphones used in the BEM calculations were measured using a laser vibrometer.

One cannot trust the diffuse-field results for LS1 microphones below 3 kHz; this is due to the time-selective procedure applied for separating the free-field and diffuse responses, and more specifically because of the roll-off frequencies of the passband filter used in the time-selective procedure. This is a step that could easily be improved by designing a filter with different roll-off frequencies. Above the lower roll-off frequency the agreement between the three estimates is within 0.1 dB below the resonance frequency (about 8 kHz for LS1 microphones and 18 kHz for LS2 microphones). Around and above the resonance frequency the agreement is still reasonably good, about 0.3 dB, but some systematic deviations occur, especially in the case of LS2 microphones. The BEM calculations agree slightly better with the diffuse-field measurements than with the random-incidence measurements. Introducing small, systematic errors in the angle of incidence in the BEM calculations indicates that the random-incidence sensitivities at high frequencies may have been underestimated because of a systematic error in the incidence angle in the anechoic room, and this is a likely explanation for the observed deviations. Nevertheless, the overall agreement is fair, and the good agreement confirms that the diffuse-field and the random-incidence corrections are equivalent when the definition of the diffuse-field sensitivity in Ref. 16 is interpreted as an ensemble average of realizations of reverberant interference fields.

Figures 8 and 9 also show the spatial standard deviations of the diffuse-field sensitivities (in decibels) and values predicted from the expression

$$\sigma\{20\log(M_d)\} \approx 5\log e\sqrt{\frac{3}{BT_{60}\log e}} \approx \frac{5.7}{\sqrt{BT_{60}}}. \qquad (7)$$

Equation (7) follows from an expression derived by Lubman[19] and Schroeder[27] for the spatial variance of the mean-square pressure averaged over a frequency band combined with the fact that the diffuse-field sensitivity is proportional to the fourth root of the frequency averaged mean-square pressure; cf. Eq. (3). The measured standard deviations are in good agreement with the predictions.

VI. CONCLUSIONS

The definition of the diffuse-field sensitivity given in Ref. 16 leads to dramatic fluctuations with the frequency and must be combined with spatial averaging and frequency smoothing. The diffuse-field sensitivities of LS1 and LS2 microphones have been determined in a small reverberation

chamber using the reciprocity technique. The experimental procedure requires (i) measurement of the electrical transfer impedance between two microphones, (ii) a time-selective procedure to separate the direct sound field from the reverberant sound field, and (iii) averaging of the resulting diffuse-field sensitivity both spatially and in the frequency domain. The resulting diffuse-field sensitivity has been found to agree satisfactorily with the random-incidence sensitivity and with the results of BEM calculations based on measured membrane velocity distributions.

The main advantage of determining the diffuse-field sensitivity using reciprocity compared with determining the random-incidence correction is that the former provides a direct, absolute estimate whereas the random-incidence correction is determined from a relative measurement. It is also worth mentioning that measurement of the random-incidence sensitivity requires facilities that are not always available, such as an anechoic room, and such measurements can be very time consuming if a high accuracy is wanted.

ACKNOWLEDGMENTS

This project has been carried out with support from the Danish Research Council for Technology and Production Sciences (FTP).

[1]F. Jacobsen and T. Roisin, "The coherence of reverberant sound fields," J. Acoust. Soc. Am. **108**, 204–210 (2000).
[2]L. L. Beranek, *Acoustical Measurements* (AIP/ASA, New York, 1988).
[3]*Microphone Handbook* (Brüel & Kjær, Nærum, Denmark, 1996).
[4]W. R. MacLean, "Absolute measurement of sound without a primary standard," J. Acoust. Soc. Am. **12**, 140–146 (1940).
[5]R. K. Cook, "Absolute pressure calibration of microphones," J. Acoust. Soc. Am. **12**, 415–420 (1941).
[6]A. L. DiMattia and F. M. Wiener, "On the absolute pressure calibration of condenser microphones by the reciprocity method," J. Acoust. Soc. Am. **18**, 341–344 (1946).
[7]W. Wathen-Dunn, "On the reciprocity free-field calibration of microphones," J. Acoust. Soc. Am. **21**, 542–546 (1949).
[8]IEC International Standard 61094-2, *Measurement microphones Part 2: Primary method for pressure calibration of laboratory standard microphones by the reciprocity technique* (1995).
[9]IEC International Standard 61094-3, *Measurement microphone Part 3: Primary method for free-field calibration of laboratory standard microphones by the reciprocity technique* (1995).
[10]S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "A time-selective technique for free-field reciprocity calibration of condenser microphones," J. Acoust. Soc. Am. **114**, 1467–1476 (2003).
[11]H. G. Diestel, "Reciprocity calibration of microphones in a diffuse sound field," J. Acoust. Soc. Am. **33**, 514–518 (1961).
[12]T. Nakajima, "Reciprocity calibration of laboratory standard microphones in a diffuse sound field," Researches of the Electrotechnical Laboratory **76**, 1–87 (1970). (in Japanese.)
[13]M. Vorländer and H. Bietz, "Novel broad-band reciprocity technique for simultaneous free-field and diffuse-field microphone calibration," Acustica **80**, 365–377 (1994).
[14]IEC International Standard 61183, *Random incidence and diffuse-field calibration of sound level meters* (1994).
[15]S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "On experimental determination of the random-incidence response," J. Acoust. Soc. Am. **121**, 2628–2636 (2007).
[16]IEC International Standard 61094-1, *Measurement microphones Part 1: Specifications for laboratory standard microphones* (1992).
[17]H. G. Diestel, "Probability distribution of sinusoidal sound pressure in a room," J. Acoust. Soc. Am. **35**, 2019–2022 (1963).
[18]R. V. Waterhouse, "Statistical properties of reverberant sound fields," J. Acoust. Soc. Am. **43**, 1436–1444 (1968).
[19]D. Lubman, "Fluctuations of sound with position in a reverberant room," J. Acoust. Soc. Am. **44**, 1491–1502 (1968).
[20]F. Jacobsen, "The diffuse sound field," Report No. 27, The Acoustics Laboratory, Technical University of Denmark, 1979.
[21]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (AIP/ASA, New York, 1989).
[22]M. R. Schroeder, "Die statistischen parameter der frequenzkurven von grossen räumen (statistical parameters of frequency responses in large rooms)," Acustica **4**, 594–600 (1954).
[23]M. R. Schroeder, "Frequency-correlation of frequency responses in rooms," J. Acoust. Soc. Am. **34**, 1819–1823 (1963).
[24]S. Müller and P. Massarani, "Transfer-function measurements with sweeps," J. Audio Eng. Soc. **49**, 443–471 (2001).
[25]R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, Jr., "Measurement of correlation coefficients in reverberant sound fields," J. Acoust. Soc. Am. **27**, 1072–1077 (1955).
[26]D. Lubman, "Spatial averaging in a diffuse sound field," J. Acoust. Soc. Am. **46**, 532–534 (1969).
[27]M. R. Schroeder, "Effect of frequency and space averaging on the transmission response of multimode media," J. Acoust. Soc. Am. **46**, 277–283 (1969).

Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation

Jean-François Durand and Christian Soize[a)]
Laboratoire Modélisation et Simulation Multi échelle, Université Paris-Est, FRE3160 CNRS, 5 Boulevard Descartes, 77454 Marne-la-Vallée, France

Laurent Gagliardini
Acoustic Division, PSA Peugeot-Citroën, 78943 Vélizy-Villacoublay, France

(Received 25 September 2007; revised 5 February 2008; accepted 8 June 2008)

The design of cars is mainly based on the use of computational models to analyze structural vibrations and internal acoustic levels. Considering the very high complexity of such structural-acoustic systems, and in order to improve the robustness of such computational structural-acoustic models, both model uncertainties and data uncertainties must be taken into account. In this context, a probabilistic approach of uncertainties is implemented in an adapted computational structural-acoustic model. The two main problems are the experimental identification of the parameters controlling the uncertainty levels and the experimental validation. Relevant experiments have especially been developed for this research in order to constitute an experimental database devoted to structural vibrations and internal acoustic pressures. This database is used to perform the experimental identification of the probability model parameters and to validate the stochastic computational model. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2953316]

PACS number(s): 43.40.At, 43.40.Hb, 43.40.Qi, 43.40.Sk [JGM] Pages: 1513–1525

I. INTRODUCTION

In the automotive industry, computational structural-acoustic models are nowadays intensively used to analyze the structural-acoustic behavior of vehicles in terms of structural vibrations and internal acoustic levels mainly for the low-frequency range. The present evolution is to extend such computational models to the medium-frequency range. In this paper, we are interested in the booming noise consisting in studying the acoustic response at passengers' ears induced by engine structure-borne excitations in the low-frequency band but also in the lower part of the medium-frequency band. Note that a very few papers have been published concerning the booming noise prediction in this frequency band using a computational model with or without experimental comparisons (Sol and Van-Herpe, 2001; Sung and Nefske, 2001; Hamdi *et al.*, 2005; Hayashi *et al.*, 2000). In addition, structural-acoustic analyses of cars in the high-frequency band are of a great importance for automotive engineering and can generally be treated by using statistical energy analysis and diffuse field methods for which numerous papers have been published in the past decade (see, for instance, Lyon and DeJong, 1995; Le Bot, 2002; Gagliardini *et al.*, 2005; Shorter and Langley, 2005; Langley, 2007). Considering the very high complexity of such structural-acoustic systems, the mean computational structural-acoustic models do not allow sufficiently good predictions to be obtained and consequently must be improved in implementing a model of uncertainties in order to increase the robustness of predictions. In this context, it is necessary to validate such computational structural-acoustic models with experiments. It should be noted that a very few complete and documented experimental databases are available in the literature. Only some elements concerning two databases can be found in the literature (Wood and Joachim, 1987; Kompella and Bernhard, 1996). Nevertheless, these two experimental databases cannot easily be used because there are no available computational structural-acoustic models associated with these databases. In order to get round this difficulty, a complete experimental campaign devoted to structural vibrations and internal acoustic pressures has been done for this research (Durand, 2007). This experimental database is presented in this paper and is used to perform an experimental identification and to validate the computational model. The problem related to such predictions with computational structural-acoustic models is extremely difficult due to the oversensitiveness of structural dynamical responses with respect to manufacturing processes and due to the small variabilities induced by the presence of optional extra around a main configuration. Note that automotive engineering requires to predict the structural-acoustic responses of a car of the same type with optional extra using only one computational structural-acoustic model. This oversensitiveness may be seen by computation when applying design changes, but also experimentally when monitoring vehicle dispersions (Wood and Joachim, 1987; Kompella and Bernhard, 1996; Hills *et al.*, 2004). Concerning the computational structural-acoustic models, many uncertainties are introduced by the mechanical-acoustical-mathematical modeling process due to the high complexity of the structure and of the internal acoustic cavity in terms of geometry, boundary conditions, material properties, etc. Even if a sophisticated structural-

[a)]Electronic mail: christian.soize@univ-paris-est.fr

acoustic model is developed, model uncertainties and data uncertainties are inherent in such a computational structural-acoustic model. One objective of this paper is to use a representative computational structural-acoustic model developed by an automotive industry (Sol and Van Herpe, 2001; Durand *et al.*, 2005a). In such a finite element model, (1) the structure is discretized with a reasonable fine scale, (2) the internal acoustic cavity is discretized with a coarse scale adapted to the frequency band of analysis, and (3) the sound-proofing schemes located at the interface between the structure and the acoustic cavity are taken into account by a very simplified model. Clearly, the sound-proofing schemes could be discretized with a fine scale using the finite element method and formulations for pore elastic materials (Attala *et al.*, 2001; Hamdi *et al.*, 2005). Such an approach has not been retained in this research. It seems that there is no paper available in the literature devoted to the development of a computational structural-acoustic model for cars including a model of uncertainties and associated with an experimental validation performed with a complete experimental database. In this paper, we present such a complete computational model including uncertainty modeling, experimental identification, and experimental validation (Durand *et al.*, 2004, 2005a and 2005b; Durand, 2007). It is known that several approaches can be used to take into account uncertainties in computational models of complex structural-acoustic systems for the low-frequency band (interval method, fuzzy set approach, probabilistic approach, etc.). In this paper, we have chosen to use the most efficient mathematical tool adapted to model uncertainties as soon as the probability theory can be used, i.e., the probabilistic approach. The mean computational structural-acoustic model is constructed from the designed system (conceptual system) using a mathematical-mechanical-acoustical modeling process. The mean computational model which is considered as a predictive model of the real system depends on parameters (or data). There are two types of uncertainties which are data uncertainties and model uncertainties. Data uncertainties are related to the parameters of the mean computational vibracoustic model and model uncertainties are induced by the modeling process. The "parametric probabilistic approach" is the most efficient and powerful method to address "data uncertainties" in predictive models (see, for instance, Schueller, 1997 and 2007, and Ghanem and Spanos, 2003, for stochastic finite element methods) as soon as the probability theory can be used but cannot address "model uncertainties." The "nonparametric probabilistic approach" recently proposed (Soize, 2000, 2001, 2003, 2005a, and 2005b) is a way to address both model uncertainties and data uncertainties. The use of the parametric probabilistic approach of data uncertainties for structural-acoustic analysis of cars generally requires to introduce a very large number of random variables. This is due to the fact that the structural-acoustic responses are very sensitive to many parameters related to the geometry (such as plate and shell thicknesses, panels curvatures, etc.), to the spot welding points, to the boundary conditions, etc. Typically, several ten thousands parameters must be modeled by random variables. First, it should be noted that the construction of the probabilistic model of this large number of parameters is not so easy to carry out. The experimental identification of a very large number of probability distributions using measurements of the response of a structural-acoustic system (solving an inverse stochastic problem and consequently, solving an optimization problem) is completely unrealistic. In addition, as explained above, the parametric probabilistic approach does not allow model uncertainties to be taken into account and it is known that model uncertainties are significant in computational structural-acoustic models of cars. This is a reason why we propose to use the nonparametric probabilistic approach of uncertainties which allows data uncertainties but also model uncertainties to be taken into account. In addition, the nonparametric approach introduces a very small number of parameters (typically seven parameters) which controls the level of uncertainties. In this condition, the experimental identification of these parameters is realistic and can be performed by solving the stochastic inverse problem using adapted mathematical-statistical tools. In this paper, we present such an approach. It should be noted that the nonparametric probabilistic approach of uncertainties has been used in the past five years for linear and nonlinear structural dynamical problems. Nevertheless, this approach has not yet been used for complex structural-acoustic systems and is presented in this paper.

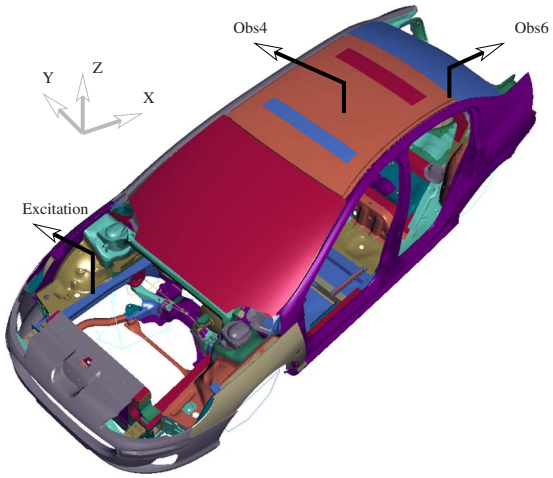


FIG. 1. (Color online) Car body, structural excitation, and structural observations.

In Sec. II, we present the experimental database which has especially been constructed for this research. Section III deals with the mean computational structural-acoustic model and Sec. IV is devoted to the stochastic reduced computational model. Finally, in Sec. V, we present the structural-acoustic response of the vehicle for which the experimental database has been carried out, the experimental identification of the probabilistic model of uncertainties and the experimental validation of the stochastic computational model.

II. EXPERIMENTS

Experiments which are described in this section have been performed in PSA Peugeot-Citroën facilities. The system under consideration is a given vehicle for which two sets of experiments have been defined. The first set is made up of

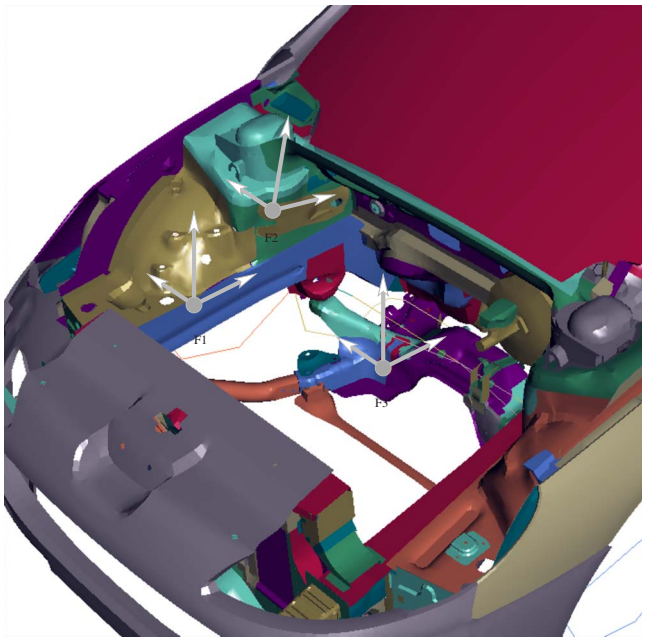


FIG. 2. (Color online) Front of the car body and structural excitations for the internal acoustic response.

structural vibration and structural-acoustic measurements. The second set is devoted to the acoustic measurements inside the internal acoustic cavity.

A. Structural-vibration and structural-acoustic measurements

The first set of experiments consists (1) of measurements of the structural frequency response functions (FRFs) between one degree of freedom (DOF) for a given excitation force (vertical component of the force at one support of the engine) and normal accelerations to the structure at six given points and (2) of measurements of the structural-acoustic FRFs between nine DOFs of excitation corresponding to the three given point forces at supports of the engine and the acoustic pressure at the driver ears located inside the internal acoustic cavity. For structural-vibration measurements, the excitation is performed by means of a hammer and the responses are identified with accelerometers. For the structural-acoustic measurements, the reciprocity method in acoustics is used. This means that the excitation is produced by an acoustic source located at the driver ears inside the internal acoustic cavity and the acceleration responses are measured at the nine DOFs introduced above. The frequency band of analysis is $[20,220]$ Hz. The experimental database has been constructed using 20 cars of the same type with different optional extras. The measurements have been realized at the exit of the assembly plant. Figure 1 shows the car body, the structural driving point (vertical excitation), and the two structural observation points (denoted by Obs4 and Obs6) for which we present experimental results, and for which comparisons with the computational structural-acoustic model is carried out. Figure 2 shows the front of the car body and the three components of the three forces applied to the points shown on this figure and corresponding to the nine DOFs of excitation. Figures 3 and 4 display the experimental structural FRFs introduced above for observations Obs4 and Obs6 and for the 20 cars. The left figures display the moduli of the FRFs in decibel and the right figures display the corresponding experimental coherence functions. It should be noted that the coherence is not really good in the frequency band $[20,70]$ Hz and consequently, the level of the moduli of the experimental FRFs contains experimental errors. This experimental difficulty is mainly induced by the hammer technique used for the excitation when the direct method is used to perform the experimental identification of the structural FRFs. Note that this technique is not used to perform the experimental identification of the structural-acoustic FRFs which are identified with the reciprocal acoustic method and for which the experimental identification is correct in the frequency band $[20,70]$ Hz. With the reciprocal acoustic method, we have not obtained any difficulties with respect to the coherence functions and then the measurements are validated in the frequency band $[20,220]$ Hz. The strategy used is then the following. The experimental structural FRFs is only used in the frequency band $[70,220]$ Hz in order to identify the dispersion parameters of the nonparametric

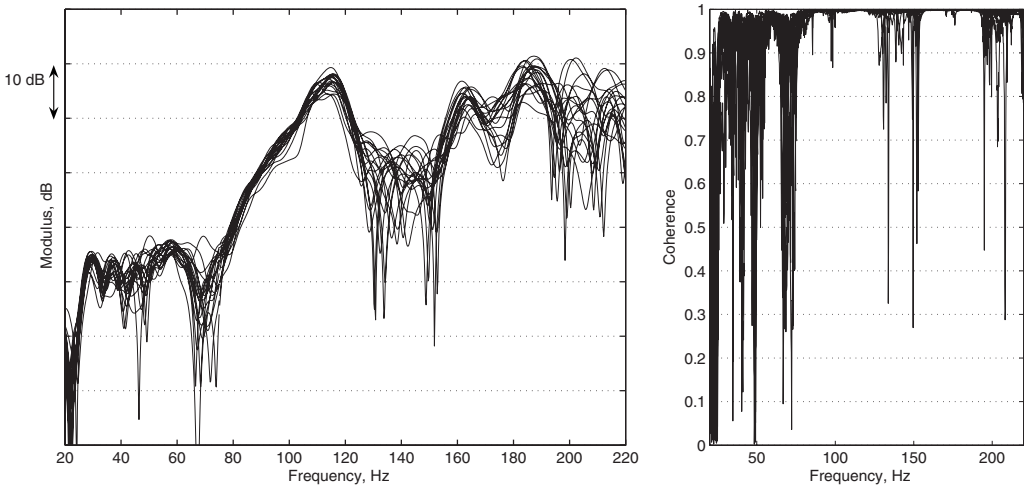


FIG. 3. Observation Obs4: Graphs of the modulus in dB scale of the experimental structural FRFs as a function of the frequency in Hz for the 20 cars (left figure). Graphs of the corresponding experimental coherence functions (right figure).

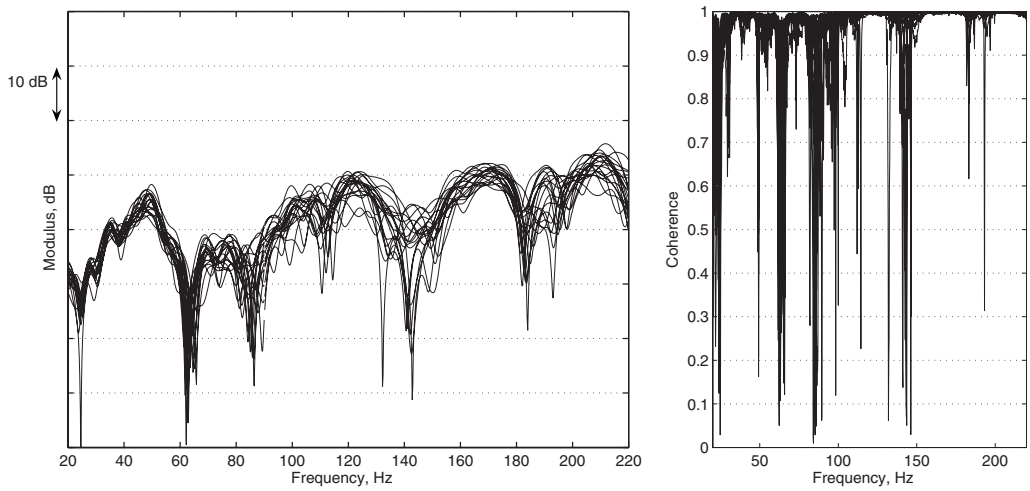


FIG. 4. Observation Obs6: Graphs of the modulus in dB scale of the experimental structural FRFs as a function of the frequency in Hz for the 20 cars (left figure). Graphs of the corresponding experimental coherence functions (right figure).

probabilistic model of uncertainties in the structure (and consequently is not used in the frequency band $[20,70]$ Hz). The experimental structural-acoustic FRFs identified with the reciprocal acoustic method is used to validate the computational structural-acoustic model in all the frequency band $[20,220]$ Hz including the frequency band $[50,183]$ Hz of the booming noise. Figures 3 and 4 show the experimental dispersion on the structural responses obtained for the 20 cars of the same type. This experimental dispersion is due to the variability induced by the optional extras and induced by the manufacturing process. This dispersion increases with the frequency which is in coherence with the expected results. The experimental dispersion varies between 5 and 10 dB in the regions close to the resonances and varies from 15 to 30 dB in the neighborhood of the antiresonances. The experimental results concerning the structural-acoustic FRFs between the nine DOFs of the three given excitation forces and the acoustic pressure at the driver ears are not presented here in order to limit the number of figures. Nevertheless, these experimental results are directly used to construct the experimental booming noise synthesis which is presented below.

B. Acoustic measurements inside the internal acoustic cavity

The second set of experiments consists of measurements of the acoustic FRFs for the acoustic pressures at 32 points inside the internal acoustic cavity whose locations are approximatively uniformly distributed inside the cavity and for an excitation which is an acoustic source located in the neighborhood of the front-passenger feet. This experimental database has been constructed for 30 different configurations of the acoustic cavity of a same car (type and position of seats, air temperature, wall boundaries between the trunk and the passenger compartment, number of passengers, etc). The frequency band of analysis is $[20,320]$ Hz. In this case and for each configuration, the observation is the root mean square $\mathrm{dB}_{\mathrm{rms}}^{\mathrm{exp}}(f)$ in decibel scale of the acoustic pressures averaged on the 32 microphones as a function of the frequency f in hertz. The reference pressure is $P_{\mathrm{ref}}=2.10^{-5}$ Pa. We then have

$$\mathrm{dB}_{\mathrm{rms}}^{\mathrm{exp}}(f) = 20\log_{10}\left(\frac{\sqrt{1/32\Sigma_{\ell=1}^{32}|p_{\ell}^{\mathrm{exp}}(2\pi f)|^2}}{P_{\mathrm{ref}}}\right),$$

in which $p_1^{\mathrm{exp}}(\omega),\ldots,p_{32}^{\mathrm{exp}}(\omega)$ are the acoustic pressures measured by the 32 microphones at frequency ω in rad/s. Figure 5 displays the graphs of $f\mapsto \mathrm{dB}_{\mathrm{rms}}^{\mathrm{exp}}(f)$ for the 30 configurations of the internal acoustic cavity. This figure shows a low level of the experimental dispersion for this quantity with respect to the different configurations. The levels of dispersion vary from 2 to 3 db. It should be noted that the moduli of the corresponding experimental FRFs at a given point (which are not presented here) have about 3 db of dispersion in the regions close to the resonances and have up to 6 dB of dispersion in the neighborhood of antiresonances.

C. Experimental synthesis of booming noise

As explained above, the measurements presented in Secs. II A and II B will be used to identify the probability

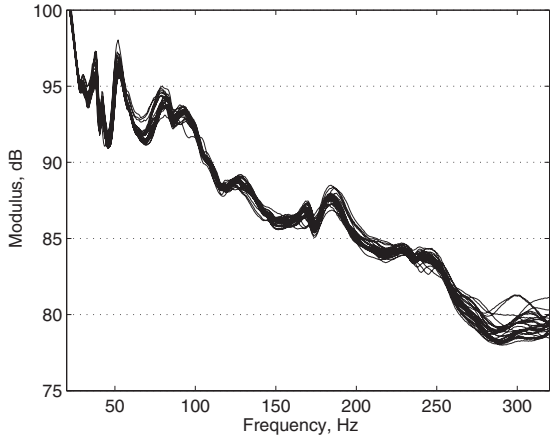


FIG. 5. For the 30 configurations of the internal acoustic cavity: graphs of $f\mapsto \mathrm{dB}_{\mathrm{rms}}^{\mathrm{exp}}(f)$.

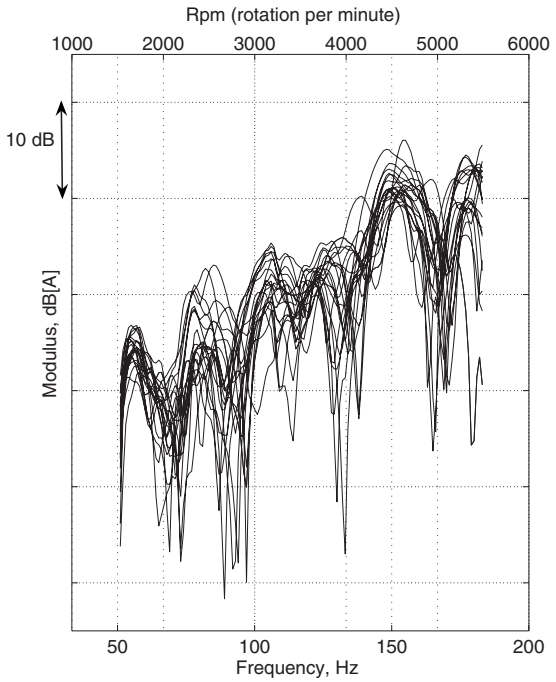


FIG. 6. Graph of $f \mapsto \mathrm{dB}^{\mathrm{exp}}[A](f)$ related to the experimental synthesized booming noise.

model of uncertainties. Below we present the measurements that are used to validate the computational structural-acoustic model allowing the booming noise to be predicted in the frequency range $[50,183]$ Hz corresponding to $[1500,5500]$ rpm (rotation per minute of the engine). In this paper, the booming noise is defined as the modulus of the acoustic pressure in dB$[A]$ at the driver ears for structural excitation induced by the forces delivered by the engine at its four supports when the frequency of rotation varies in the frequency band of analysis. This function is denoted by $f \mapsto \mathrm{dB}^{\mathrm{exp}}[A](f)$. Note that the forces delivered by the engine at its supports correspond to the second harmonic of rotation (for instance, 1500 rpm corresponds to a frequency equal to 50 Hz). The scale used is the dB$[A]$ scale which corresponds to the weighting of the usual decibel scale by the A-weighting. Figure 6 displays the graph of $f \mapsto \mathrm{dB}^{\mathrm{exp}}[A](f)$ related to the experimental synthesized booming noise and defined above. In this figure, a very large experimental dispersion induced by the variability (optional extras) and by the manufacturing process can be seen. As for the experimental structural FRFs, the experimental dispersion varies between 5 and 10 dB in the regions close to the resonances and varies from 15 to 30 dB in the neighborhood of the antiresonances. These experimental results are in coherence with the experimental structural FRFs and can be explained by the fact that the experimental dispersion induced by uncertainties inside the internal acoustic cavity is relatively small with respect to structural uncertainties (see Fig. 5).

III. REDUCED MEAN COMPUTATIONAL STRUCTURAL-ACOUSTIC MODEL

The mean computational structural-acoustic model is developed in the context of three dimension linear elastoacoustics, is formulated in the frequency domain, and is discretized by using the finite element method. The structural-acoustic system is made up of a damped elastic structure coupled with a closed internal acoustic cavity filled with a dissipative acoustic fluid. The external acoustic fluid in air and its effects on the structural-acoustic system are assumed to be negligible in the low-frequency band of analysis which is considered in the paper and which is devoted to the booming noise prediction. The linear responses of the structural-acoustic system are studied around a static equilibrium state which is taken as the natural state at rest. The reduced mean computational structural-acoustic model is derived from the mean model using a modal analysis.

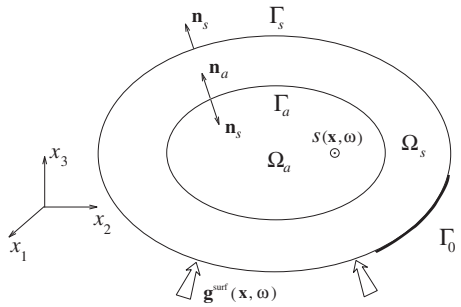


FIG. 7. Scheme of the structural-acoustic system.

A. Description of the structural-acoustic system

We then consider linear vibrations of a fixed damped structure Ω_s (the car body) subjected to external loads (engine excitations), coupled with its internal cavity Ω_a (passenger compartment and trunk) filled with a dissipative acoustic fluid (air). We are mainly interested in predicting the frequency responses of the structural-acoustic system in the frequency band of analysis $B=[\omega_{\min},\omega_{\max}]$ rad/s corresponding to $B_f=[f_{\min},f_{\max}]$ Hz with $\omega_{\min}=2\pi f_{\min}$ and $\omega_{\max}=2\pi f_{\max}$. The physical space $\mathbb{R}^3$ is referred to a Cartesian system and the generic point of $\mathbb{R}^3$ is denoted by $\mathbf{x}=(x_1,x_2,x_3)$. The system is defined in Fig. 7. Let $\partial\Omega_s=\Gamma_s\cup\Gamma_0\cup\Gamma_a$ be the boundary of Ω_s. The outward unit normal to $\partial\Omega_s$ is denoted by $\mathbf{n}_s=(n_{s,1},n_{s,2},n_{s,3})$. The displacement field in Ω_s is denoted by $\mathbf{u}(\mathbf{x},\omega)=(u_1(\mathbf{x},\omega),u_2(\mathbf{x},\omega),u_3(\mathbf{x},\omega))$. The structure is assumed to be fixed on the part Γ_0 of the boundary $\partial\Omega_s$. The internal acoustic cavity Ω_a is the bounded domain filled with a dissipative acoustic fluid. The boundary $\partial\Omega_a$ of Ω_a is Γ. The outward unit normal to $\partial\Omega_a$ is denoted by $\mathbf{n}_a=(n_{a,1},n_{a,2},n_{a,3})$ and we then have $\mathbf{n}_a=-\mathbf{n}_s$ on $\partial\Omega_a$. The pressure field in Ω_a is denoted by $p(\mathbf{x},\omega)$.

B. Mean boundary value problem for the structural-acoustic system

The equation of the structural part is written (e.g., Trusdell, 1960; Ohayon and Soize, 1998) as

$$-\omega^2\rho_s u_i-\frac{\partial\sigma_{ij}}{\partial x_j}=g_i^{\mathrm{vol}} \quad \text{in } \Omega_s, \tag{1}$$

with the convention for summations over repeated Latin indices, in which ρ_s is the mass density, σ_{ij} is the stress tensor, $\mathbf{u}=(u_1,u_2,u_3)$ is the displacement field of the structure, and

$\mathbf{g}^{\text{vol}}=(g_1^{\text{vol}},g_2^{\text{vol}},g_3^{\text{vol}})$ is the body force field applied to the structure. The boundary conditions are

$$\sigma_{ij}(\mathbf{u})n_{s,j}=g_i^{\text{surf}} \quad \text{on } \Gamma_s, \quad \sigma_{ij}(\mathbf{u})n_{s,j}=-pn_{s,i} \quad \text{on } \Gamma_a,$$

$$\mathbf{u}=0 \quad \text{on } \Gamma_0, \tag{2}$$

in which $\mathbf{g}^{\text{surf}}=(g_1^{\text{surf}},g_2^{\text{surf}},g_3^{\text{surf}})$ is the surface force field applied to the structure. The damped structure is made up with a linear nonhomogeneous anisotropic viscoelastic material without memory. In the frequency domain, the constitutive equation is written as $\sigma_{ij}=a_{ijkh}\epsilon_{kh}+i\omega b_{ijkh}\epsilon_{kh}$, in which the tensor a_{ijkh} of the elastic coefficients and the tensor b_{ijkh} of the damping coefficients of the material depend on $\mathbf{x}$, are independent of ω, and have the usual properties of symmetry and positive definiteness. The strain tensor ϵ_{kh} is related to displacement field $\mathbf{u}$ by $\epsilon_{kh}=(\partial u_k/\partial x_h+\partial u_h/\partial x_k)/2$. Concerning the internal dissipative acoustic fluid, we use the $\mathbf{u}-p$ formulation. The equation governing the vibration of the dissipative acoustic fluid occupying domain Ω_a is written as (Ohayon and Soize, 1998; Pierce, 1989; Lighthill, 1978)

$$\frac{\omega^2}{\rho_a c_a^2}p+i\omega\frac{\tau}{\rho_a}\nabla^2 p+\frac{1}{\rho_a}\nabla^2 p=\frac{\tau}{\rho_a}c_a^2\nabla^2 s-i\frac{\omega}{\rho_a}s \quad \text{in } \Omega_a, \tag{3}$$

for which the boundary conditions are

$$\frac{1}{\rho_a}(1+i\omega\tau)\frac{\partial p}{\partial \mathbf{n}_a}=\omega^2\mathbf{u}\cdot\mathbf{n}_a+\tau\frac{c_a^2}{\rho_a}\frac{\partial s}{\partial \mathbf{n}_a} \quad \text{on } \Gamma_a, \tag{4}$$

where ρ_a is the mass density of the acoustic fluid at equilibrium, c_a is the speed of sound, τ is the coefficient due to the viscosity of the fluid (the coefficient due to thermal conduction is neglected), and $s(\mathbf{x},\omega)$ is the acoustic source density.

C. Mean computational structural-acoustic model

The finite element method (Zienkiewicz and Taylor, 2000) is used to solve numerically the above boundary value problem. We consider a finite element mesh of the structure Ω_s and of the internal acoustic fluid Ω_a. Let $\underline{\mathbf{u}}^s=(\underline{u}_1^s,\ldots,\underline{u}_{n_s}^s)$ be the complex vector of the n_s DOFs of the structure according to the finite element discretization of the displacement field $\mathbf{u}$. Let $\underline{\mathbf{p}}^a=(\underline{p}_1^a,\ldots,\underline{p}_{n_a}^a)$ be the complex vector of the n_a DOFs of the acoustic fluid to the finite element discretization of the pressure field p. The finite element discretization of the boundary value problem in terms of $\mathbf{u}$ and p (Ohayon and Soize, 1998), defined by Eqs. (1)–(4) with the Dirichlet condition on Γ_0 yields the mean computational structural-acoustic model,

$$\begin{bmatrix} [\underline{A}_{n_s}^s(\omega)] & [\underline{C}_{n_s,n_a}] \\ \omega^2[\underline{C}_{n_s,n_a}]^T & [\underline{A}_{n_a}^a(\omega)] \end{bmatrix} \begin{bmatrix} \underline{\mathbf{u}}^s(\omega) \\ \underline{\mathbf{p}}^a(\omega) \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{f}}^s(\omega) \\ \underline{\mathbf{f}}^a(\omega) \end{bmatrix}, \tag{5}$$

where $[\underline{A}_{n_s}^s(\omega)]$ is the dynamical stiffness matrix of the damped structure *in vacuo* which is a symmetric $(n_s\times n_s)$ complex matrix such that

$$[\underline{A}_{n_s}^s(\omega)]=-\omega^2[\underline{M}_{n_s}^s]+i\omega[\underline{D}_{n_s}^s]+[\underline{K}_{n_s}^s],$$

in which $[\underline{M}_{n_s}^s]$, $[\underline{D}_{n_s}^s]$, and $[\underline{K}_{n_s}^s]$ are the mass, damping, and stiffness matrices of the structure which are positive-definite symmetric $(n_s\times n_s)$ real matrices. In Eq. (5), $[\underline{A}_{n_a}^a(\omega)]$ is the dynamical stiffness matrix of the dissipative acoustic fluid which is a symmetric $(n_a\times n_a)$ complex matrix such that

$$[\underline{A}_{n_a}^a(\omega)]=-\omega^2[\underline{M}_{n_a}^a]+i\omega[\underline{D}_{n_a}^a]+[\underline{K}_{n_a}^a],$$

in which $[\underline{M}_{n_a}^a]$, $[\underline{D}_{n_a}^a]$, and $[\underline{K}_{n_a}^a]$ are the "mass," "damping," and "stiffness" matrices of the acoustic cavity with fixed coupling interface. The matrix $[\underline{M}_{n_a}^a]$ is a positive-definite symmetric $(n_a\times n_a)$ real matrix, and $[\underline{D}_{n_a}^a]$ and $[\underline{K}_{n_a}^a]$ are the positive-semidefinite symmetric $(n_a\times n_a)$ real matrices whose ranks are n_a-1. The matrix $[C_{n_s,n_a}]$ is the structural-acoustic coupling matrix which is a $(n_s\times n_a)$ real matrix.

D. Reduced mean computational structural-acoustic model

The projection of the mean computational structural-acoustic model on the structural modes *in vacuo* and on the acoustic modes of the acoustic cavity with fixed coupling interface yields the reduced mean computational structural-acoustic model. The structural modes *in vacuo* and the acoustic modes of the cavity with fixed coupling interface are calculated by solving the two generalized eigenvalue problems

$$[\underline{K}_{n_s}^s]\underline{\boldsymbol{\psi}}=\underline{\lambda}^s[\underline{M}_{n_s}^s]\underline{\boldsymbol{\psi}} \tag{6}$$

$$[\underline{K}_{n_a}^a]\underline{\boldsymbol{\phi}}=\underline{\lambda}^a[\underline{M}_{n_a}^a]\underline{\boldsymbol{\phi}}. \tag{7}$$

The eigenvectors verify the usual orthogonal properties (Bathe and Wilson, 1976; Géradin and Rixen, 1994; Ohayon and Soize; 1998). The structural displacement is written as

$$\underline{\mathbf{u}}^s(\omega)=[\underline{\Psi}]\underline{\mathbf{q}}^s(\omega), \tag{8}$$

in which $[\underline{\Psi}]$ is the $(n_s\times n)$ real matrix whose columns are constituted of the n structural modes associated with the n first positive eigenvalues (the n first structural eigenfrequencies). The internal acoustic cavity has one constant pressure mode and $m-1$ acoustic modes. The internal acoustic pressure is written as

$$\underline{\mathbf{p}}^a(\omega)=[\underline{\Phi}]\underline{\mathbf{q}}^a(\omega), \tag{9}$$

in which $[\underline{\Phi}]$ is the $(n_a\times m)$ real matrix whose columns are constituted (1) of the constant pressure mode associated with the zero eigenvalue and (2) of the acoustic modes associated with the positive eigenvalues (the $m-1$ first acoustical eigenfrequencies). It should be noted that the constant pressure mode is kept in order to model the quasistatic variation of the internal fluid pressure induced by the deformation of the coupling interface (Ohayon and Soize, 1998). Using Eqs. (8) and (9), the projection of Eq. (5) yields the reduced mean matrix model of the structural-acoustic system

$$\begin{bmatrix} [\underline{\mathcal{A}}_n^s(\omega)] & [\underline{\mathcal{C}}_{n,m}] \\ \omega^2[\underline{\mathcal{C}}_{n,m}]^T & [\underline{\mathcal{A}}_m^a(\omega)] \end{bmatrix} \begin{bmatrix} \underline{\mathbf{q}}^s(\omega) \\ \underline{\mathbf{q}}^a(\omega) \end{bmatrix} = \begin{bmatrix} \underline{\mathcal{F}}^s(\omega) \\ \underline{\mathcal{F}}^a(\omega) \end{bmatrix}, \tag{10}$$

where $[\underline{\mathcal{C}}_{n,m}]=[\underline{\Psi}]^T[\underline{C}_{n_s,n_a}][\underline{\Phi}]$, $[\underline{\mathcal{F}}^s(\omega)]=[\underline{\Psi}]^T[\underline{\mathbf{f}}^s(\omega)]$, and $[\underline{\mathcal{F}}^a(\omega)]=[\underline{\Phi}]^T[\underline{\mathbf{f}}^a(\omega)]$. The generalized dynamical stiffness matrix $[\underline{\mathcal{A}}_n^s(\omega)]$ of the damped structure is written as

$$[\underline{\mathcal{A}}_n^s(\omega)] = -\omega^2[\underline{\mathcal{M}}_n^s] + i\omega[\underline{\mathcal{D}}_n^s] + [\underline{\mathcal{K}}_n^s],$$

in which $[\underline{\mathcal{M}}_n^s]=[\underline{\Psi}]^T[\underline{M}_{n_s}^s][\underline{\Psi}]$ and $[\underline{\mathcal{K}}_n^s]=[\underline{\Psi}]^T[\underline{K}_{n_s}^s][\underline{\Psi}]$ are the positive-definite diagonal $(n\times n)$ real matrices and where $[\underline{\mathcal{D}}_n^s]=[\underline{\Psi}]^T[\underline{D}_{n_s}^s][\underline{\Psi}]$ is, in general, a positive-definite full $(n\times n)$ real matrix. The generalized dynamical stiffness matrix $[\underline{\mathcal{A}}_m^a(\omega)]$ of the dissipative acoustic fluid is written as

$$[\underline{\mathcal{A}}_m^a(\omega)] = -\omega^2[\underline{\mathcal{M}}_m^a] + i\omega[\underline{\mathcal{D}}_m^a] + [\underline{\mathcal{K}}_m^a],$$

in which $[\underline{\mathcal{M}}_m^a]=[\underline{\Phi}]^T[\underline{M}_{n_a}^a][\underline{\Phi}]$ is a positive-definite diagonal $(m\times m)$ real matrix and where $[\underline{\mathcal{D}}_m^a]=[\underline{\Phi}]^T[\underline{D}_{n_a}^a][\underline{\Phi}]$ and $[\underline{\mathcal{K}}_m^a]=[\underline{\Phi}]^T[\underline{K}_{n_a}^a][\underline{\Phi}]$ are the positive-semidefinite diagonal $(m\times m)$ real matrices of rank $m-1$.

IV. STOCHASTIC REDUCED COMPUTATIONAL STRUCTURAL-ACOUSTIC MODEL

As explained in the Introduction, both data uncertainties and model uncertainties can be taken into account by using the nonparametric probabilistic approach of uncertainties. Such an approach (Soize, 2000, 2001, and 2005a) has been used in linear and nonlinear structural dynamics. Nevertheless, this approach has not yet been used for complex structural-acoustic systems which requires the use of extended results concerning random matrix theory in order to take into account model uncertainties for the structural-acoustic coupling operator. Such extended results have been recently proposed by Soize (2005b).

A. Constructing the stochastic model

The use of the nonparametric probabilistic approach (Durand, 2007) of both model uncertainties and data uncertainties for the structure, the acoustic cavity, and the structural-acoustic coupling consists (1) in modeling the generalized mass $[\underline{\mathcal{M}}_n^s]$, damping $[\underline{\mathcal{D}}_n^s]$, and stiffness $[\underline{\mathcal{K}}_n^s]$ matrices of the structure by random matrices $[\boldsymbol{M}_n^s]$, $[\boldsymbol{D}_n^s]$, and $[\boldsymbol{K}_n^s]$ whose dispersion parameters are denoted by δ_{M^s}, δ_{D^s}, and δ_{K^s}, respectively; (2) in modeling the generalized mass $[\underline{\mathcal{M}}_m^a]$, damping $[\underline{\mathcal{D}}_m^a]$, and stiffness $[\underline{\mathcal{K}}_m^a]$ matrices of the acoustic cavity by random matrices $[\boldsymbol{M}_m^a]$, $[\boldsymbol{D}_m^a]$, and $[\boldsymbol{K}_m^a]$ whose dispersion parameters are denoted by δ_{M^a}, δ_{D^a}, and δ_{K^a}, respectively; (3) in modeling the generalized structural-acoustic coupling matrix $[\underline{\mathcal{C}}_{n,m}]$ by a random matrix $[\boldsymbol{C}_{n,m}]$ whose dispersion parameter is denoted by δ_C. The explicit construction of the probability distribution of these random matrices is given by (Soize 2000, 2001, and 2005b; Durand, 2007) for random matrices $[\boldsymbol{M}_n^s]$, $[\boldsymbol{D}_n^s]$, $[\boldsymbol{K}_n^s]$, $[\boldsymbol{M}_m^a]$, $[\boldsymbol{D}_m^a]$, $[\boldsymbol{K}_m^a]$, and $[\boldsymbol{C}]$. Let $[\boldsymbol{H}]$ be any of these random matrices. The probability distribution of such a random matrix $[\boldsymbol{H}]$ depends on its mean value $[\underline{H}]=E\{[\boldsymbol{H}]\}$, where E is the mathematical expectation and depends on its dispersion parameter δ_H which must be taken independent of the matrix dimension. An algebraic representation of random matrix $[\boldsymbol{H}]$ has been developed and allows independent realizations to be constructed for a stochastic solver based on the Monte Carlo numerical simulation. This algebraic representation is recalled below. For random matrices $[\boldsymbol{M}_n^s]$, $[\boldsymbol{D}_n^s]$, $[\boldsymbol{K}_n^s]$, $[\boldsymbol{M}_m^a]$, $[\boldsymbol{D}_m^a]$, and $[\boldsymbol{K}_m^a]$, random matrix $[\boldsymbol{H}]$ is then a symmetric positive-definite (or positive-semidefinite) real-valued random matrix and $[\boldsymbol{H}]$ is written as $[\boldsymbol{H}]=[\underline{L}_H]^T[\boldsymbol{G}_H][\underline{L}_H]$, in which $[\underline{H}]=[\underline{L}_H]^T[\underline{L}_H]$ and where $[\boldsymbol{G}_H]$ is the random matrix germ. Note that if $[\underline{H}]$ is positive semidefinite then the factorization is not a Cholesky factorization and is obtained by another algebraic algorithm. When $[\boldsymbol{H}]$ is the rectangular matrix $[\boldsymbol{C}_{n,m}]$ (see Soize, 2005b), using the following polar decomposition $[\underline{\mathcal{C}}_{n,m}]=[\underline{U}][\underline{T}]$ with $[\underline{U}]^T[\underline{U}]=[I]$, where $[\underline{T}]$ is a positive-definite matrix which can then be factorized as $[\underline{T}]=[\underline{L}_C]^T[\underline{L}_C]$, the random matrix $[\boldsymbol{C}_{n,m}]$ can then be written as $[\boldsymbol{C}_{n,m}]=[\underline{U}][\underline{L}_C]^T[\boldsymbol{G}_H][\underline{L}_C]$ in which $[\boldsymbol{G}_H]$ is another random matrix germ. The stochastic reduced model of the uncertain structural-acoustic system for which the reduced mean model is defined by Eqs. (8)–(10) is written for all ω fixed in the frequency band of analysis as

$$\boldsymbol{U}^s(\omega) = [\underline{\Psi}]\boldsymbol{Q}^s(\omega), \quad \boldsymbol{P}^a(\omega) = [\underline{\Phi}]\boldsymbol{Q}^a(\omega), \tag{11}$$

in which the $\mathbb{C}^n$-valued random variable $\boldsymbol{Q}^s(\omega)=(Q_1^s(\omega),\ldots,Q_n^s(\omega))$ and the $\mathbb{C}^m$-valued random variable $\boldsymbol{Q}^a(\omega)=(Q_1^a(\omega),\ldots,Q_m^a(\omega))$ are the solution of the following random matrix equation:

$$\begin{bmatrix} [\mathbf{A}_n^s(\omega)] & [\mathbf{C}_{n,m}] \\ \omega^2[\mathbf{C}_{n,m}]^T & [\mathbf{A}_m^a(\omega)] \end{bmatrix} \begin{bmatrix} \boldsymbol{Q}^s(\omega) \\ \boldsymbol{Q}^a(\omega) \end{bmatrix} = \begin{bmatrix} \underline{\mathcal{F}}^s(\omega) \\ \underline{\mathcal{F}}^a(\omega) \end{bmatrix}, \tag{12}$$

in which the random complex matrices $[\boldsymbol{A}_n^s(\omega)]$ and $[\boldsymbol{A}_m^a(\omega)]$ are defined by $[\boldsymbol{A}_n^s(\omega)]=-\omega^2[\boldsymbol{M}_n^s]+i\omega[\boldsymbol{D}_n^s]+[\boldsymbol{K}_n^s]$ and where $[\boldsymbol{A}_m^a(\omega)]=-\omega^2[\boldsymbol{M}_m^a]+i\omega[\boldsymbol{D}_m^a]+[\boldsymbol{K}_m^a]$. Let $\boldsymbol{G}_H$ be the positive-definite symmetric $(\nu\times\nu)$ random matrix representing $[\boldsymbol{G}_{M_n^s}]$, $[\boldsymbol{G}_{D_n^s}]$, $[\boldsymbol{G}_{K_n^s}]$, $[\boldsymbol{G}_{M_m^a}]$, $[\boldsymbol{G}_{D_m^a}]$, $[\boldsymbol{G}_{K_m^a}]$, or $[\boldsymbol{G}_{C_{n,m}}]$. The following algebraic representation of random matrix $[\boldsymbol{G}_H]$ allows independent realizations of $[\boldsymbol{G}_H]$ to be constructed. Random matrix $[\boldsymbol{G}_H]$ is written $[\boldsymbol{G}_H]=[\boldsymbol{L}_H]^T[\boldsymbol{L}_H]$ in which $[\boldsymbol{L}_H]$ is a random upper triangular $(\nu\times\nu)$ real matrix whose random elements are independent random variables defined as follows:

(1) For $j<j'$, the real-valued random variable $[\boldsymbol{L}_H]_{jj'}$ is written as $[\boldsymbol{L}_H]_{jj'}=\sigma_\nu U_{jj'}$ in which $\sigma_\nu=\delta_H(\nu+1)^{-1/2}$ and where $U_{jj'}$ is a real-valued Gaussian random variable with zero mean and variance equal to 1. The parameter δ_H controlling the dispersion level of random matrix $[\boldsymbol{H}]$ is such that $\delta_H=\sqrt{E\{\|[\boldsymbol{G}_H]-[I]\|_F^2\}}/\nu$ in which $[I]$ is the identity matrix and where the subindex F corresponds to the Frobenius norm.

(2) For $j=j'$, the positive-valued random variable $[\boldsymbol{L}_H]_{jj'}$ is written as $[\boldsymbol{L}_H]_{jj'}=\sigma_\nu\sqrt{2V_j}$ in which σ_ν is defined above and where V_j is a positive-valued gamma random variable whose probability density function p_{V_j} with respect to dv is written as

$$p_{V_j}(v) = \mathbb{1}_{\mathbb{R}^+}(v)\frac{1}{\Gamma\left(\nu+1/2\,\delta_H^2+1-j/2\right)}v^{(\nu+1/2\delta_H^2-1-j/2)}e^{-v}.$$

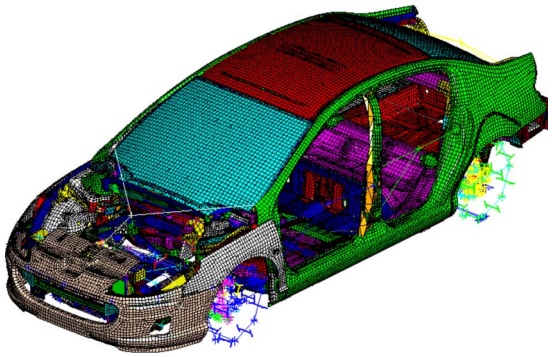


FIG. 8. (Color online) Finite element mesh of the structure: 978 733 structural DOFs.

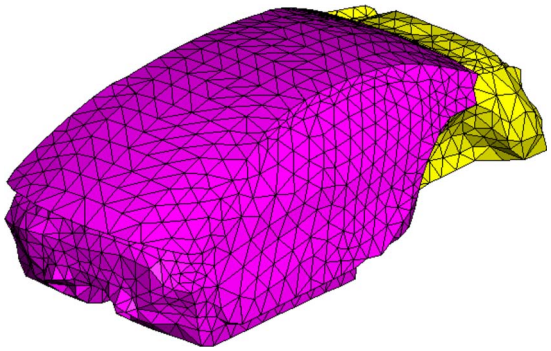


FIG. 9. (Color online) Finite element mesh of the acoustic cavity: 8139 acoustic pressure DOFs.

B. Confidence region of the random responses

Let $\omega \mapsto W(\omega)$ from B into $\mathbb{R}$ be a random observation of the structural-acoustic system in the frequency domain for which n_r independent realizations $\omega \mapsto W(\omega, \theta_1), \ldots, \omega \mapsto W(\omega, \theta_{n_r})$ are computed using the stochastic model presented in Sec. IV A with the Monte Carlo numerical simulation. The confidence region associated with the probability level α for the random function $\{W(\omega), \omega \epsilon B\}$ is constructed by using the method of quantiles (Serfling, 1980). For fixed ω in B, let $F_{W(\omega)}$ be the cumulative distribution function (continuous from the right) of random variable $W(\omega)$ which is such that $F_{W(\omega)}(w) = P(W(\omega) \leq w)$. For $0 < p < 1$, the pth quantile or fractile of $F_{W(\omega)}$ is defined as $\zeta(p) = \inf\{w : F_{W(\omega)} \times (w) \geq p\}$. Then the upper envelope $w^+(\omega)$ and the lower envelope $w^-(\omega)$ of the confidence region are defined by $w^+(\omega) = \zeta((1+\alpha)/2)$ and $w^-(\omega) = \zeta((1-\alpha)/2)$. The estimation of $w^+(\omega)$ and $w^-(\omega)$ is performed by using the sample quantiles.

V. STRUCTURAL-ACOUSTIC RESPONSE OF THE VEHICLE

In this section, we present the experimental identification of the dispersion parameters of the nonparametric probabilistic approach of uncertainties for vibroacoustic analysis of the vehicle for which the experimental database has been presented in Sec. II. We then present the validation of the structural-acoustic model including uncertainties in comparing the identified stochatic computational model with experiments (Durand, 2007).

A. Description of the mean computational structural-acoustic model and experimental comparisons

The mean computational structural-acoustic model of the vehicle is a finite element model with 978 733 structural DOFs and 8139 acoustic pressure DOFs in the internal acoustic cavity. Figure 8 displays the finite element mesh of the structure and Fig. 9 shows the finite element mesh of the acoustic cavity. Finally, Fig. 10 deals with the finite element mesh of the computational structural-acoustic model. The reduced mean computational model is constructed using $n = 1723$ elastic modes of the structure *in vacuo* and $m = 57$ acoustic modes of the internal acoustic cavity with rigid walls. These values of n and m have been deduced from a mean-square convergence analysis of the random response.

Concerning the damping of the mean computational structural-acoustics model, the generalized damping matrix of the structure and of the internal acoustic cavity are assumed to be diagonal matrices for all frequencies in B. Note that such a diagonal assumption introduces model uncertainties in the mean computational model which are taken into account by the nonparametric probabilistic approach. Consequently, the damping part in the mean model is described in terms of damping rates which have been fixed to their nominal values. As it can be seen on the experimental FRFs, the structural damping is significant and the internal acoustic cavity damping is relatively high due to the presence of absorbing materials. For the structural FRFs, Figs. 11 and 12 display the comparisons between the mean computational structural-acoustic model predictions and the experiments for the moduli of the FRFs in decibel concerning observations Obs4 (Fig. 11) and Obs6 (Fig. 12) and for the 20 cars. For the root mean square of the acoustic pressures averaged on the 32 microphones inside the internal acoustic cavity (see Sec. II B), Fig. 13 displays the graphs of $f \mapsto \mathrm{dB}_{\mathrm{rms}}^{\mathrm{exp}}(f)$ for the 30 configurations of the internal acoustic cavity and the corresponding graph $f \mapsto \mathrm{dB}_{\mathrm{rms}}(f)$ calculated with the mean computational model. Figure 14 displays the graph of $f \mapsto \mathrm{dB}^{\mathrm{exp}}[A](f)$ related to the experimental synthesized booming noise (see Sec. II C) and the corresponding graph $f \mapsto \mathrm{dB}[A](f)$ calculated with the mean computational model. Figures 11–14 show that there are significant differences between measurements and numerical simulations which are

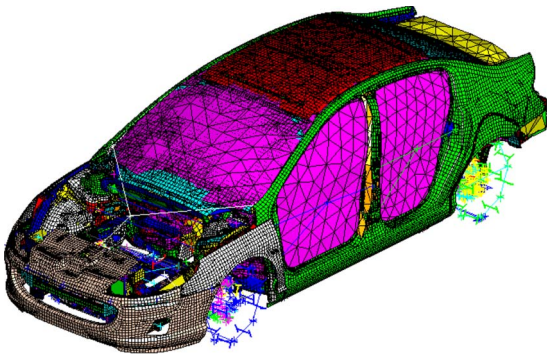


FIG. 10. (Color online) Finite element mesh of the computational structural-acoustic model.

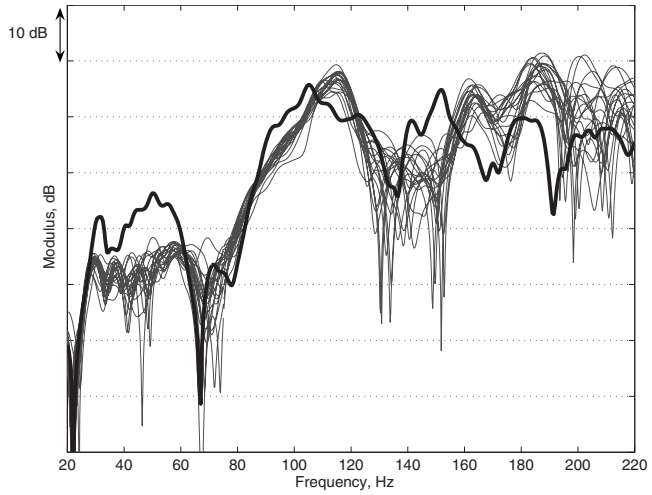


FIG. 11. Comparisons of the mean computational model results with the experiments. Graphs of the modulus in dB scale of the structural FRF for observation Obs4: experiments for the 20 cars (gray lines); mean computational model (thick solid line).

not due to the variabilities of experiments (optional extras and manufacturing process) but are mainly due to model uncertainties. As explained in Sec. I, these figures clearly show that both data uncertainties and model uncertainties have to be taken into account in the mean computational structural-acoustic model in order to improve the predictability and the robustness of the predictions.

B. Methodology and assumptions for the experimental identification of the dispersion parameters

The methodology and the assumptions used to identify the dispersion parameters δ_{M^s}, δ_{D^s}, and δ_{K^s} for the structures, δ_{M^a}, δ_{D^a}, and δ_{K^a} for the internal acoustic cavity and δ_C for the structural-acoustic coupling interface are the following.

(1) The dispersion parameters δ_{M^a}, δ_{D^a}, and δ_{K^a} of the internal acoustic cavity are identified using the experimental database defined in Sec. II B. For this identification, it is assumed that $\delta_{M^a}=\delta_{D^a}=\delta_{K^a}$. First, this hypothesis allows the computational cost to be decreased. Second, different analyses have shown that the confidence regions are not very sensitive to the value of dispersion parameter δ_{D^a}. This is due to the fact that the modal density is relatively high in the frequency band of analysis and in addition, the internal acoustic cavity has a large dissipative factor due to the presence of the absorbing materials as explained above. Finally, a sensitivity analysis of the responses has been carried out with respect to δ_{M^a} and δ_{K^a} and has shown that the confidence regions were very sensitive to the values of the dispersion parameters but that $\delta_{M^a}=\delta_{K^a}$ could be written. The method used to iden-

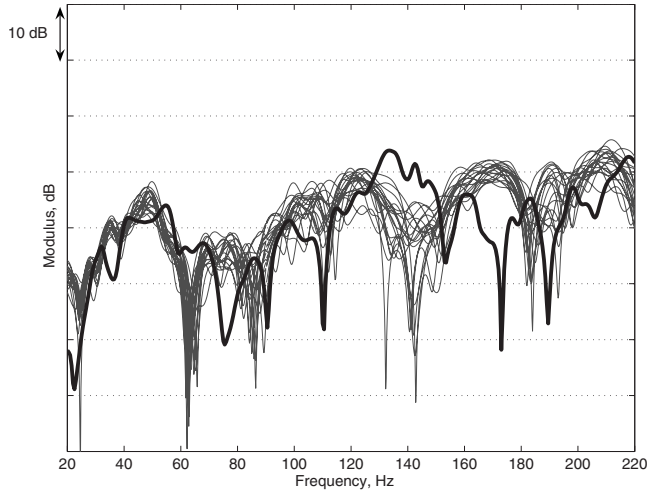


FIG. 12. Comparisons of the mean computational model results with the experiments. Graphs of the modulus in dB scale of the structural FRF for observation Obs6: experiments for the 20 cars (gray lines); mean computational model (thick solid line).

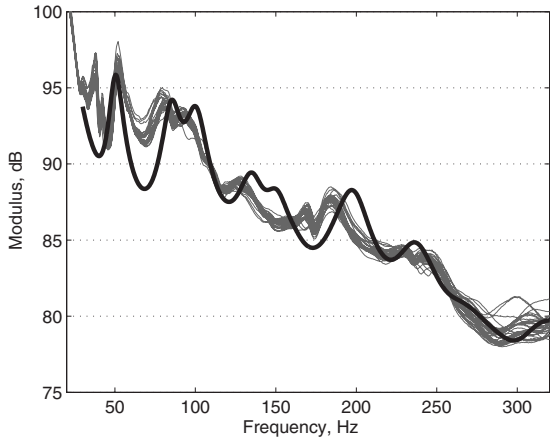


FIG. 13. Comparisons of the mean computational model results with the experiments. Graphs of the root mean square of the acoustic pressures averaged in the cavity in dB scale: experiments for the 30 configurations (gray lines); mean computational model (thick solid line).

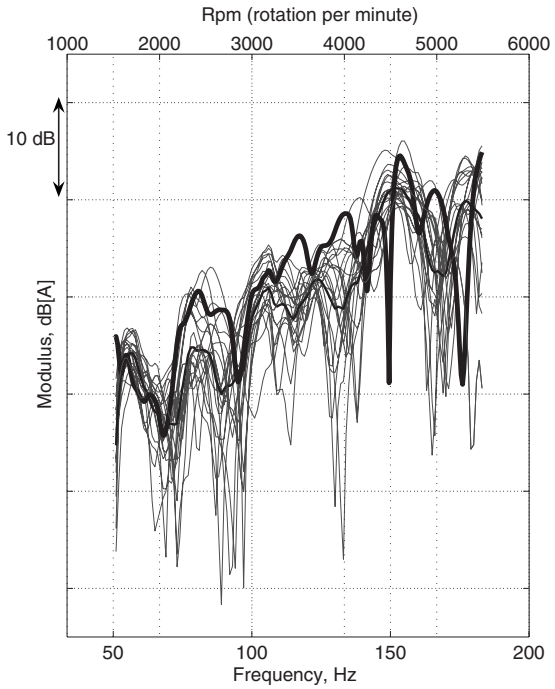


FIG. 14. Comparisons of the mean computational model results with the experiments. Graphs of the modulus of the synthesized booming noise in dB[A] scale: experiments for the 20 cars (gray lines); mean value of the experiments (thin solid line); mean computational model (thick solid line).

tify the dispersion parameter δ_a such that $\delta_a = \delta_{M^a} = \delta_{D^a} = \delta_{K^a}$ is the maximum likelihood method (Spall, 2003; Walter and Pronzato, 1997).

(2) The dispersion parameters δ_{M^s}, δ_{D^s}, and δ_{K^s} for the structure are identified using the structural FRFs (and not the structural-acoustic FRFs) of the experimental database defined in Sec. II A. Therefore, the stochastic computational structural-acoustic model must be used. Then, for the identification of these dispersion parameters, dispersion parameters δ_{M^a}, δ_{D^a}, and δ_{K^a} are fixed to the identified experimental value δ_a and the value of δ_C is fixed to a nominal value. This last hypothesis is reasonable because the confidence regions of the random structural responses are not sensitive to the acoustic coupling (note that the structural acoustic responses are sensitive to δ_C, but we are presently speaking about the identification of δ_{M^s}, δ_{D^s}, and δ_{K^s} using only the structural responses). Concerning δ_{D^s}, the situation is similar to the case of the internal acoustic cavity, but for the structure, δ_{D^s} has been fixed to a nominal value which results from previous studies. In addition, it should be noted that the confidence regions of the structural responses are a little sensitive to δ_{D^s} compared with the high sensitivity of these responses with respect to δ_{M^s} and δ_{K^s}. Concerning the experimental identification of δ_{M^s} and δ_{K^s}, the assumption $\delta_s = \delta_{M^s} = \delta_{K^s}$ (similar to the assumption used for the internal acoustic cavity) has been replaced by the following one: $\delta_s = \sqrt{\delta_{M^s}^2 + \delta_{K^s}^2}$. This choice results from computational tests and is much more efficient and more representative than the other one. For computational cost reasons (see below), the maximum likelihood method could not be used and consequently has been substituted by the mean-square method with a differentiable objective function (Spall, 2003; Walter and Pronzato, 1997).

(3) For the identification of the dispersion parameter δ_C of the structural-acoustic coupling interface, the structural-acoustic FRFs of the experimental database defined in Sec. II C are used. Therefore, the stochastic computational structural-acoustic model must be used. Then, for the identification of δ_C, dispersion parameters δ_{M^a}, δ_{D^a}, δ_{K^a}, δ_{M^s}, δ_{D^s}, and δ_{K^s} are fixed to their identified values (see above). For similar reasons to those given above, the maximum likelihood method cannot be used and should then be substituted by the mean-square method. Unfortunately, the corresponding objective function is not sufficiently sensitive to the value of δ_C and the maximum likelihood method is not efficient. So we have identified the value of δ_C using a trial method.

C. Experimental identification of the dispersion parameters of the internal acoustic cavity

As explained above, the maximum likelihood method is used to identify the dispersion parameter δ_a introduced in Sec. V B 1. The random observation Y and its corresponding experimental quantity $Y^{\exp}$ used for this identification are defined by

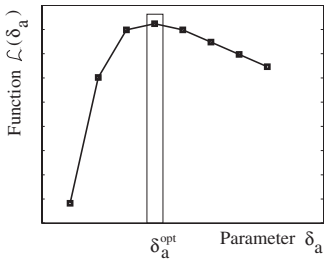


FIG. 15. Graph of $\delta_a \mapsto \mathcal{L}(\delta_a)$.

$$Y = \int_{B_f} \mathrm{dB}_{\mathrm{rms}}(f)df, \quad Y^{\exp} = \int_{B_f} \mathrm{dB}_{\mathrm{rms}}^{\exp}(f)df,$$

in which $\mathrm{dB}_{\mathrm{rms}}(f)$ corresponds to $\mathrm{dB}_{\mathrm{rms}}^{\exp}(f)$ defined in Sec. II B for the experiments and which is written as

$$\mathrm{dB}_{\mathrm{rms}}(f) = 20\log_{10}\left(\frac{\sqrt{1/32\Sigma_{\ell=1}^{32}|\boldsymbol{P}_{j_\ell}^a(2\pi f)|^2}}{\boldsymbol{P}_{\mathrm{ref}}}\right),$$

in which $j_1, \ldots, j_{32}$ are the DOFs corresponding to the measurement points. Let $Y^{\exp}(\eta_1), \ldots, Y^{\exp}(\eta_{30})$ be the independent experimental realizations corresponding to the 30 configurations of the internal acoustic cavity. Let $\delta_a \mapsto \mathcal{L}(\delta_a)$ be the log-likelihood function defined by

$$\mathcal{L}(\delta_a) = \sum_{\nu=1}^{30} \log_{10} p_Y(\delta_a, Y^{\exp}(\eta_\nu)),$$

in which $y \mapsto p_Y(\delta_a, y)$ is the probability density function of random variable Y. For all fixed δ_a, the right-hand side of the above equation is estimated by the Monte Carlo method with the stochastic computational structural-acoustic model presented in Sec. IV A. The maximum likelihood method consists in solving the following optimization problem:

$$\delta_a^{\mathrm{opt}} = \arg\max_{\delta_a} \mathcal{L}(\delta_a).$$

This method needs an accurate estimation of $p_Y(\delta_a, Y^{\exp}(\eta_\nu))$ which has been estimated with 20 000 independent realizations for the Monte Carlo simulation. Instead of using an optimization algorithm to solve the above optimization problem, the values of the log-likelihood function have directly calculated in eight values of δ_a in order to construct an approximation of its graph on the interval of interest. Figure 15 displays the approximation of the graph of $\delta_a \mapsto \mathcal{L}(\delta_a)$ allowing the optimal value δ_a^{opt} to be estimated. For this optimal value of δ_a^{opt}, Fig. 16 compares the experimental measurements with the computational results for the root mean square $\mathrm{dB}_{\mathrm{rms}}^{\exp}(f)$ of the acoustic pressures averaged on the 32 microphones inside the internal acoustic cavity. In this figure, (1) the 30 gray lines represent the experimental measurements, (2) the upper and lower thick solid lines represent the upper and lower envelopes of the confidence region calculated for a probability level of 0.96, (3) the mid thin solid line represents the mean value of the random response of the stochastic reduced computational model, and (4) the dashed line represents the response of the reduced mean computational model. It can be seen that the experiments belongs to the confidence region with a probability level of 0.96 that

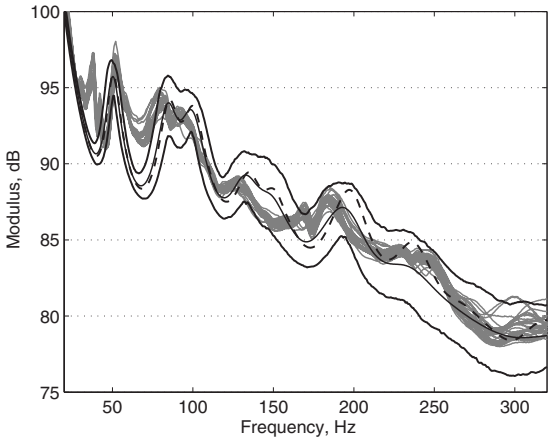


FIG. 16. Comparisons of the stochastic computational model results with the experiments. Graphs of the root mean square of the acoustic pressures averaged in the cavity in dB scale: experiments for the 30 configurations (gray lines); mean computational model (dashed line); mean value of the random response (mid thin solid line); confidence region: the upper and lower envelopes are the upper and lower thick solid lines.

validates the acoustic part of the stochatic computational model.

D. Experimental identification of the dispersion parameters of the structure

As explained in Sec. V B 2, the mean-square method with differentiable objective function is used. The dispersion parameters δ_{M^a}, δ_{D^a}, δ_{K^a}, δ_{D^s}, and δ_C are fixed to the values identified above (see Secs. V B 1 and V B 2). Consequently, we have to solve the optimization problem defined by

$$\delta_s^{\text{opt}} = \arg\min_{\delta_s} J(\delta_s),$$

in which the objective function $J(\delta_s)$ depends only on δ_s and is defined by

$$J(\delta_s) = 2(1-\gamma)\|\|\boldsymbol{Z}(\delta_s) - \boldsymbol{m}(\delta_s)\|\|^2 + 2\gamma\|\boldsymbol{m}(\delta_s) - \underline{\boldsymbol{Z}}^{\text{exp}}\|_B^2,$$

with $\gamma=0.5$. In this equation, the random observation $\boldsymbol{Z}(\omega,\delta_s)=(Z_1(\omega,\delta_s),\ldots,Z_6(\omega,\delta_s))$ of the stochastic computational model related to the six structural observations (see Sec. II A) is such that

$$Z_j(\omega,\delta_s) = \log_{10}(\omega^2|\boldsymbol{U}_{k_j}^s(\omega,\delta_s)|),$$

and its mean value $\boldsymbol{m}(\omega,\delta_s)=(m_1(\omega,\delta_s),\ldots,m_6(\omega,\delta_s))$ is

$$\boldsymbol{m}(\omega,\delta_s) = E\{\boldsymbol{Z}(\omega,\delta_s)\}.$$

The corresponding experimental observation is denoted by $\boldsymbol{Z}^{\text{exp}}(\omega)=(Z_1^{\text{exp}}(\omega),\ldots,Z_6^{\text{exp}}(\omega))$ and its mean value is

$$\underline{Z}_j^{\text{exp}}(\omega) = \frac{1}{20}\sum_{\ell=1}^{20} Z_j^{\text{exp}}(\omega,\eta_\ell).$$

Finally, the norms are defined by

$$\|\|\boldsymbol{Z}(\delta_s) - \boldsymbol{m}(\delta_s)\|\|^2 = E\left\{\sum_{j=1}^{6}\int_B |Z_j(\omega,\delta_s) - m_j(\omega,\delta_s)|^2 d\omega\right\}$$

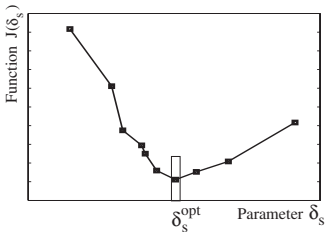


FIG. 17. Graph of $\delta_s \mapsto J(\delta_s)$.

$$\|\boldsymbol{m}(\delta_s) - \underline{\boldsymbol{Z}}^{\text{exp}}\|_B^2 = \sum_{j=1}^{6}\int_B |m_j(\omega,\delta_s) - \underline{Z}_j^{\text{exp}}(\omega)|^2 d\omega.$$

The first norm represents the variance of the computational model due to the uncertainties and the second norm represents the bias between the experiments and the stochastic model. In this objective function, $B=2\pi B_f$ with $B_f=[100,180]$ Hz and the frequency resolution is 0.5 Hz. For each evaluation of the objective function, the stochastic reduced computational model is solved using the Monte Carlo method with 1000 independent realizations and corresponding to a mean-square convergence of the second-order stochastic solution. Since each evaluation of the objective function requires about 500 h of CPU time (the computations have been realized with 20 CPU yielding an elapsed time of 25 h for each evaluation of the objective function), we have used a trial method consisting in computing the cost function for ten values of the couple $(\delta_{M_s},\delta_{K_s})$. It can then be deduced the approximation of the graph of $\delta_s \mapsto J(\delta_s)$ for the ten values of $\delta_s=\sqrt{\delta_{M_s}^2+\delta_{K_s}^2}$. Figure 17 displays this approximation of the graph of $\delta_s \mapsto J(\delta_s)$, allowing the optimal value δ_s^{opt} to be estimated. From δ_s^{opt}, the optimal values $\delta_{M_s}^{\text{opt}}$ and $\delta_{K_s}^{\text{opt}}$ of δ_{M_s} and δ_{K_s} can then be deduced. For these optimal values $\delta_{M_s}^{\text{opt}}$ and $\delta_{K_s}^{\text{opt}}$ of δ_{M_s} and δ_{K_s}, and for the values of the other dispersion parameters fixed to their values previously identified, Figs. 18 and 19 compare the experimental measure-

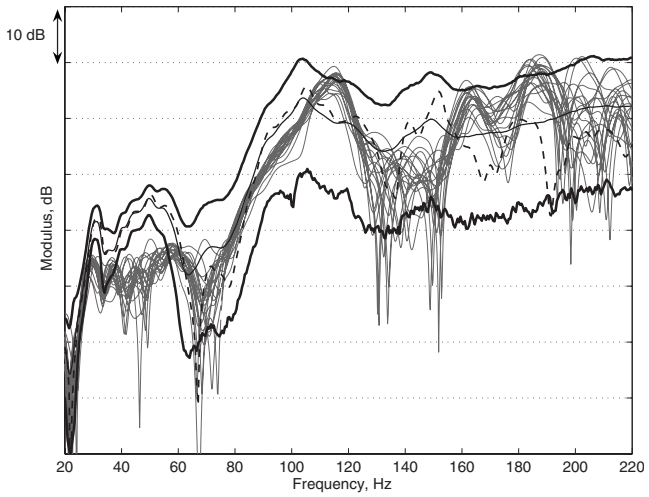


FIG. 18. Comparisons of the stochastic computational model results with the experiments for observation Obs4. Graphs of the moduli of the FRFs in dB scale: experiments for the 20 cars (gray lines); mean computational model (dashed line); mean value of the random response (mid thin solid line); confidence region: the upper and lower envelopes are the upper and lower thick solid lines.

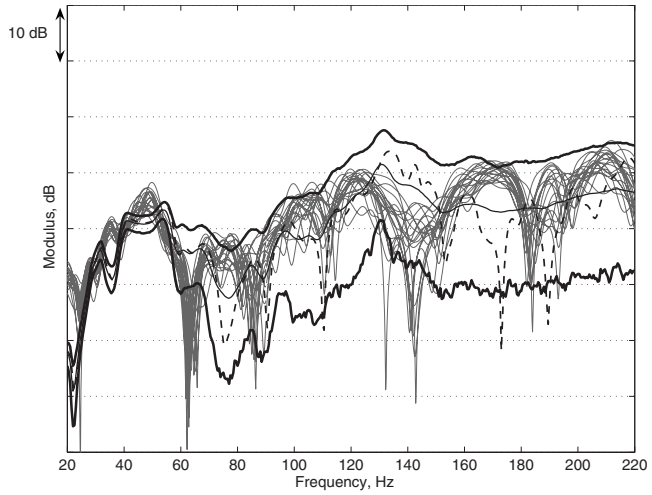


FIG. 19. Comparisons of the stochastic computational model results with the experiments for observation Obs6. Graphs of the moduli of the FRFs in dB scale: experiments for the 20 cars (gray lines); mean computational model (dashed line); mean value of the random response (mid thin solid line); confidence region: the upper and lower envelopes are the upper and lower thick solid lines.

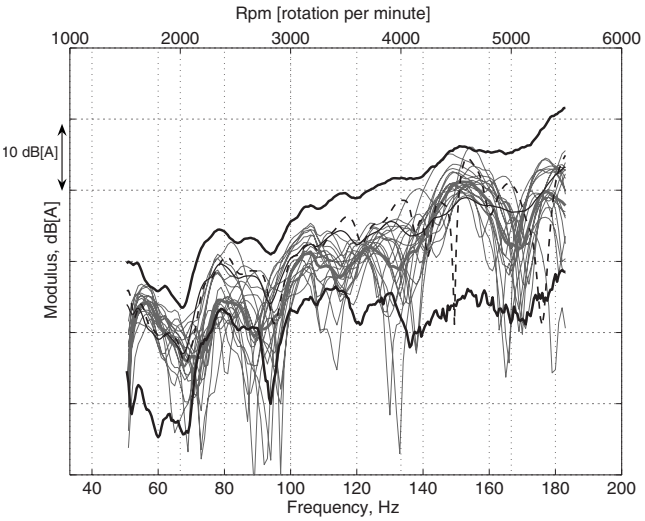


FIG. 20. Comparisons of the stochastic computational model results with the experiments for the booming noise. Graphs of the moduli of the FRFs in dB[A] scale: experiments for the 20 cars (thin gray lines); mean value of the experiments (thick gray line); mean computational model (dashed line); mean value of the random response (mid thin solid line); confidence region: the upper and lower envelopes are the upper and lower thick solid lines.

ments with the computational results for the moduli of the FRFs in decibel concerning observations Obs4 (Fig. 18) and Obs6 (Fig. 19) and for the 20 cars. In these figures, (1) the 20 gray lines represent the experimental measurements, (2) the upper and lower thick solid lines represent the upper and lower envelopes of the confidence region calculated for a probability level of 0.96, (3) the mid thin solid line represents the mean value of the random response of the stochastic reduced computational model, and (4) the dashed line represents the response of the reduced mean computational model. It can be seen that the experiments belong to the confidence region with a probability level of 0.96 that validates the structural part of the stochatic computational model.

E. Experimental validation of the stochastic computational structural-acoustic model for the booming noise

In this section, we compare the experiments for the synthesized booming noise of the database (see Sec. II C) with the predictions given by the stochastic computational structural-acoustic model. For the calculation, the dispersion parameters δ_{M^a}, δ_{D^a}, δ_{K^a}, δ_{M^s}, δ_{D^s}, δ_{K^s}, and δ_C are fixed to their optimal values which are either fixed or experimentally identified as explained in the previous sections. The Monte Carlo stochastic solver is used with 1500 realizations. A mean-square convergence analysis of the stochastic response has been carried out (Durand, 2007) and the convergence is reached for this number of realizations. Figure 20 compares the experimental measurements with the computational results for the booming noise in dB[A] scale. The 20 thin gray lines represent the experimental measurements of the booming noise for the 20 cars and the thick gray line the mean value of the experiments. The upper and lower thick solid lines represent the upper and lower envelopes of the confidence region calculated for a probability level of 0.96. The mid thin solid line represents the mean value of the random response of the stochastic reduced computational model. The mid dashed line represents the response of the reduced mean computational model. Taking into account the complexity of the vibroacoustic model, the obtained results validate the stochastic computational model and demonstrate its capability to predict experimental measurements knowing that the dispersion parameters of model uncertainties have been identified.

VI. CONCLUSION

In this paper, we have presented and validated a methodology to analyze a very complex structural-acoustic system in the low- and medium-frequency ranges. Both data uncertainties and model uncertainties have been taken into account in the computational model. The nonparametric probabilistic approach introduces a small number of dispersion parameters which can then be experimentally identified in solving optimization problems. In the context of the automative industry, a large experimental database has been constructed and has been presented in this paper. This database has been used for the identification of the probabilistic model. An experimental validation has been proposed for each step of the identification. The global experimental validation of the stochastic computational model has been obtained. The comparisons are good enough taking into account the complexity of the system and taking into account that the mean computational model is unique to represent all the variabilities induced by optional extras and manufacturing process. In addition, this mean computational model has voluntary not been updated with the measurements (this is the practical situation encountered by engineering in such an automative industry). With respect to the real structural-acoustic system, the dynamical behavior of the soundproofing schemes (insulation materials) existing in the real

cars (and in particular in the cars for which the experimental database has been constructed) has not been taken into account in the mean computational model (but has been taken into account as pure mass subsystems). Such a model could be improved in including such a dynamical model of the sound-proofing schemes.

Attala, N., Hamdi, M. A., and Panneton, R. (**2001**). "Enhanced weak integral formulation for the mixed (u,p) poroelastic equations," J. Acoust. Soc. Am. **109**, 3065–3068.

Bathe, K. J., and Wilson, E. L. (**1976**). *Numerical Methods in Finite Element Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

Durand, J. F., Gagliardini, L., and Soize, C. (**2004**). "Random modeling of frequency response functions of cars," International Conference on Modal Analysis Noise and Vibration Engineering (ISMA 2004) Proceedings.

Durand, J. F., Gagliardini, L., and Soize, C. (**2005a**). "Nonparametric modeling of the variability of vehicle structural-acoustic behavior," SAE 2005 Noise and Vibration Conference Proceedings.

Durand, J. F., Gagliardini, L., and Soize, C. (**2005b**). "Nonparametric modeling of structural-acoustic coupling uncertainties," EuroDyn 2005 Conference Proceedings.

Durand, J. F. (**2007**). "Structural-acoustic modeling of automotive vehicles in presence of uncertainties and experimental identification and validation," Ph.D. thesis, Université de Marne-la-Vallée, France, 2007.

Gagliardini, L., Houillon, L., Borello, G., and Petrinelli, L. (**2005**). "Virtual SEA-FEA-based modeling of mid-frequency structure borne noise," Sound Vib. **39**, 22–28.

Géradin, M., and Rixen, D. (**1994**). *Mechanical Vibrations*, Wiley, Chichester, UK.

Ghanem, R., and Spanos, P. D. (**2003**). *Stochastic Finite Elements: A Spectral Approach*, revised ed., Dover, New York.

Hamdi, M. A., Zhang, C., Mebarek, L., Anciant, M., and Mathieux, B. (**2005**). "Analysis of structural-acoustic performances of a fully trimmed vehicle using an innovative subsystem solving approach faciliting the cooperation between carmakers and sound-package suppliers," EuroDyn 2005 Conference Proceedings.

Hayashi, K., Yamaguchi, S., and Matsuda, A. (**2000**). "Analysis of booming noise in light-duty truck cab," J. of the Soc. of Automot. Eng. of Japan **21**, 255–257.

Hills, E., Mace, B., and Ferguson, N. S. (**2004**). "Statistics of complex built-up structures," International Conference on Modal Analysis Noise and Vibration Engineering (ISMA 2004) Proceedings.

Kompella, M. S., and Bernhard, R. J. (**1996**). "Variation of structural-acoustic characteristics of automotive vehicles," Noise Control Eng. J. **44**, 93–99.

Langley, R. S. (**2007**). "On the diffuse field reciprocity relationship and vibrational energy variance in a random subsystem at high frequencies," J. Acoust. Soc. Am. **121**, 913–921.

LeBot, A. (**2002**). "Energy transfer for high frequencies in built-up structures," J. Sound Vib. **250**, 247–275.

Lighthill, J. (**1978**). *Waves in Fluids*, Cambridge University, Cambridge.

Lyon, R. H., and Dejong, R. G. (**1995**). *Theory and applications of statistical energy analysis*, edited by Butterworth-Heinemann, Boston.

Ohayon, R., and Soize, C. (**1998**). *Structural Acoustics and Vibration*, Academic, San Diego.

Pierce, A. D. (**1989**). *Acoustics: An Introduction to its Physical Principles and Applications* (Acoust. Soc. Am., Woodbury, NY).

Schueller, G. I. (**1997**). "A state-of-the-art report on computational stochastic mechanics," Probab. Eng. Mech. **12**, 197–321.

Schueller, G. I. (**2007**). "On the treatment of uncertainties in structural mechanics and analysis," Comput. Struct. **85**, 235–243.

Serfling, R. J. (**1980**). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shorter, P. J., and Langley, R. S. (**2005**). "Vibro-acoustic analysis of complex systems," J. Sound Vib. **288**, 669–699.

Soize, C. (**2000**). "A nonparametric model of random uncertainties for reduced matrix models in structural dynamics," Probab. Eng. Mech. **15**, 277–294.

Soize, C. (**2001**). "Maximum entropy approach for modeling random uncertainties in transient elastodynamics," J. Acoust. Soc. Am. **109**, 1979–1996.

Soize, C. (**2003**). "Random matrix theory and non-parametric model of random uncertainties in vibration analysis," J. Sound Vib. **263**, 893–916.

Soize, C. (**2005a**). "A comprehensive overview of a non-parametric probablistic approach of model uncertainties for predictive models in structural dynamics," J. Sound Vib. **288**, 623–652.

Soize, C. (**2005b**). "Random matrix theory for modeling random uncertainties in computational mechanics," Comput. Methods Appl. Mech. Eng. **194**, 1333–1366.

Sol, A., and Van-Herpe, F. (**2001**). "Numerical prediction of a whole car vibroacoustic behavior at low frequencies," SAE 2001 Noise and Vibration Conference Proceedings.

Spall, J. C. (**2003**). *Introduction to Stochastic Search and Optimization*, Wiley, Hoboken, NJ.

Sung, H. S., and Nefske, D. J. (**2001**). "Assessment of a vehicle concept finite element model for predicting structural vibration," SAE 2001 Noise and Vibration Conference Proceedings.

Trusdell, C. (**1960**). *The Elements of Continuum Mechanics* (Springer-Verlag, Berlin).

Walter, E., and Pronzato, L. (**1997**). *Identification of Parametric Models from Experimental Data* (Springer, New York).

Wood, L. A., and Joachim, C. A. (**1987**). "Interior noise scatter in four-cylinder sedans and wagons," Int. J. Veh. Des. **8**, 428–438.

Zienkiewicz, O. C., and Taylor, R. L. (**2000**). *The Finite Element Method*, 5th ed. (Butterworth-Heinemann, Oxford), Vols. **1–3**.

A 2D fast near-field method for calculating near-field pressures generated by apodized rectangular pistons

Duo Chen[a] and Robert J. McGough[b]
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824

(Received 5 November 2007; revised 28 May 2008; accepted 30 May 2008)

Analytical two-dimensional (2D) integral expressions are derived for fast calculations of time-harmonic and transient near-field pressures generated by apodized rectangular pistons. These 2D expressions represent an extension of the fast near-field method (FNM) for uniformly excited pistons. After subdividing the rectangular piston into smaller rectangles, the pressure produced by each of the smaller rectangles is calculated using the uniformly excited FNM expression for a rectangular piston, and the total pressure generated by an apodized rectangular piston is the superposition of the pressures produced by all of the subdivided rectangles. By exchanging summation variables and performing integration by parts, a 2D apodized FNM expression is obtained, and the resulting expression eliminates the numerical singularities that are otherwise present in numerical models of pressure fields generated by apodized rectangular pistons. A simplified time space decomposition method is also described, and this method further reduces the computation time for transient pressure fields. The results are compared with the Rayleigh–Sommerfeld integral and the FIELD II program for a rectangular source with each side equal to four wavelengths. For time-harmonic calculations with a 0.1 normalized root mean square error (NRMSE), the apodized FNM is 4.14 times faster than the Rayleigh–Sommerfeld integral and 59.43 times faster than the FIELD II program, and for a 0.01 NRMSE, the apodized FNM is 12.50 times faster than the Rayleigh–Sommerfeld integral and 155.06 times faster than the FIELD II program. For transient calculations with a 0.1 NRMSE, the apodized FNM is 2.31 times faster than the Rayleigh–Sommerfeld integral and 4.66 times faster than the FIELD II program, and for a 0.01 NRMSE, the apodized FNM is 11.90 times faster than the Rayleigh–Sommerfeld integral and 24.04 times faster than the FIELD II program. Thus, the 2D apodized FNM is ideal for fast pressure calculations and for accurate reference calculations in the near-field region.
 [DOI: 10.1121/1.2950081]

PACS number(s): 43.40.Rj, 43.20.Px, 43.20.El, 43.35.Bf [TDM] Pages: 1526–1537

I. INTRODUCTION

Modeling the near-field pressure generated by a spatially varying particle velocity on the face of a rectangular piston is of practical importance for many acoustics applications[1,2] where accurate and fast computer simulations are required. Fast ultrasound simulations are especially important for calculating pressure fields in large computational domains, specifically steady-state High Intensity Focused Ultrasound (HIFU) simulations for ultrasound therapy[3,4] and transient calculations for ultrasound imaging.[5,6] Individual array elements are typically modeled as single baffled rectangular sources with uniform surface particle velocities, but the particle velocity on the transducer face is in general nonuniform. For example, Lin *et al.*[7] showed that the surface particle velocity of a fluid-loaded piezoelectric element on a phased array is nonuniform, and Prego Borges *et al.*[8] demonstrated that the surface velocity of a single array element is apodized and delayed when matching layers are used.

Pressure fields from uniform rectangular sources can be calculated using several numerical methods including the impulse response method,[9–11] the FIELD II program,[12,13] the Rayleigh–Sommerfeld integral,[14] and the fast near-field method (FNM).[15] Among these methods, the FNM achieves the smallest error in the least time.[15–17] The FNM for uniformly excited rectangular pistons[15] and circular pistons[17] eliminates the $1/R$ singularity and therefore avoids problems with large errors in the near-field region. The FNM also converges rapidly in the near field, which has been demonstrated for both time-harmonic and transient excitations.[15,17,18] For uniform excitations, the FNM is ideal for near-field pressure calculations and for reference calculations that evaluate the numerical errors associated with other methods.[19,20]

Although several methods, including the Rayleigh–Sommerfeld integral[14] and the FIELD II program,[12,13] calculate the pressures generated by apodized rectangular pistons, the numerical performance of these methods suffers in the near-field region. The numerical evaluation of the Rayleigh–Sommerfeld integral converges slowly in the near-field region because of the singularity introduced by the $1/R$ term, which approaches infinity when R approaches zero. The FIELD II program subdivides a rectangular piston into smaller rectangles and computes the pressure using the far-field approximation for the impulse response of the velocity potential, which also contains a $1/R$ term. Thus, the FIELD II pro-

[a]Electronic mail: chenduo@msu.edu
[b]Electronic mail: mcgough@egr.msu.edu

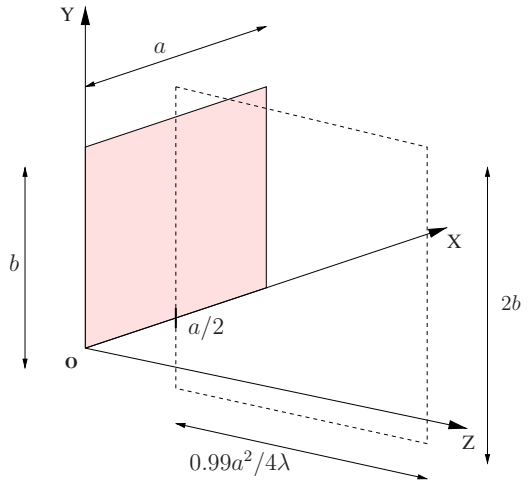


FIG. 1. (Color online) Orientation of the computational grid relative to the rectangular source. The rectangular source, which has width a and height b, lies in the $z=0$ plane. The dashed lines define the extent of the computational grid in the $x=a/2$ plane. The extent of the computational grid is $2b$ by $0.99a^2/4\lambda$ in the y and z directions, respectively.

gram generates relatively large errors and converges slowly in the near-field region, especially near the piston face. To address this problem for circular pistons, the FNM has been recently extended to include axisymmetric particle velocity distributions, and the resulting two-dimensional (2D) integral also demonstrates rapid convergence.[21] However, this apodized FNM expression is specific to pistons with circular or cylindrical symmetry, and methods for modeling apodized rectangular pistons are still needed.

To improve the performance of near-field calculations for apodized rectangular pistons, an apodized FNM expression is derived from the FNM expression for uniformly excited rectangular pistons. The derivation of the apodized FNM for rectangular pistons begins by subdividing the piston into small uniformly excited subelements. The total pressure is obtained by superposing by the pressure produced by all of the subelements. After performing a summation variable exchange and integrating by parts, the apodized FNM expression for rectangular pistons is obtained. Next, the apodized FNM for transient pressure calculations is obtained by inverse Fourier transforming the time-harmonic apodized FNM expression. The apodized FNM expression, the Rayleigh–Sommerfeld integral, and the FIELD II program are then evaluated in the near field of a square piston that extends four wavelengths in both directions. The results of time-harmonic and transient computations indicate that, when compared to calculations performed with the Rayleigh–Sommerfeld integral and the FIELD II program, (1) the apodized FNM achieves the smallest errors for a given amount of time and (2) the apodized FNM requires the least time to achieve a given error.

II. METHODS

In the derivation that follows, each observation point is denoted by (x,y,z), and each source point is denoted by $(\mu,\nu,0)$. Figure 1 shows the coordinate system used in the derivation and subsequent evaluations. The rectangular source is located in the $z=0$ plane, and the origin of the coordinate system O coincides with the lower left corner of the rectangular source. The apodization function of the source is given by $f(\mu,\nu)$, which is zero outside of the rectangular source. The width of the rectangular source is a and the height is b.

The FNM expression for a rectangular piston that is excited uniformly is given by McGough.[15] A more general unapodized FNM expression for a rectangular piston is obtained from the expression in Ref. 16 for a triangular piston. Here, the FNM expression for the uniformly excited rectangular piston is denoted by $p_0[\mu,\nu](x,y,z;k)$, where the subscript "0" indicates the uniform excitation and μ and ν represent the width and height of the rectangular source, respectively. The near-field pressure for the uniformly excited piston is given by

$$p_0[\mu,\nu](x,y,z;k) = -\frac{\rho c v_0 e^{j\omega t}}{2\pi} \times \sum_{i=1}^{4} \int_{n_i}^{m_i} h_i \frac{e^{-jk\sqrt{\sigma^2+z^2+h_i^2}} - e^{-jkz}}{\sigma^2+h_i^2} d\sigma, \tag{1}$$

where ω is the excitation frequency in rad/s, ρ is the density of the medium, c is the speed of sound, v_0 is the constant normal particle velocity evaluated on the surface of the rectangular source, k is the wave number, and σ is an integration variable as defined in Refs. 15 and 16. The values of m_i and n_i, which are functions of μ and ν, represent the upper limits and lower limits of the integral, respectively. In Eq. (1), the values of (m_i,n_i) are $(m_1,n_1)=(m_2,n_2)=(\mu-x,-x)$ and $(m_3,n_3)=(m_4,n_4)=(\nu-y,-y)$, and the values of h_i are $h_1=y$, $h_2=\nu-y$, $h_3=x$, and $h_4=\mu-x$ for $i=1,2,3,4$, respectively. According to Eq. (1), there are two special cases, namely,

$$p_0[0,\nu](x,y,z;k)=0 \text{ and } p_0[\mu,0](x,y,z;k)=0, \tag{2}$$

where the subelement has zero width or height, respectively. These special cases are utilized in the following derivation of the apodized FNM expression for a rectangular source.

A. Steady-state apodized FNM expression

The pressure field $p_{\text{apod}}(x,y,z;k)$ is obtained by subdividing the rectangular source into $N\times N$ small rectangles, where the subscript apod indicates that the pressure is computed with the apodization function $f(\mu,\nu)$. One of these small rectangles is depicted inside of the rectangular source in Fig. 2. The values $\Delta\mu=a/N$ and $\Delta\nu=b/N$ denote the width and the height of each small rectangle in the X and Y directions, respectively, and $S[i,j]$ represents the rectangle at the ith column and jth row of the subdivision. The four coordinates of the vertices of $S[i,j]$ are given by $((i-1)\Delta\mu,(j-1)\Delta\mu)$, $(i\Delta\mu,(j-1)\Delta\mu)$, $(i\Delta\mu,j\Delta\mu)$, and $((i-1)\Delta\mu,j\Delta\mu)$ for the lower left, lower right, upper right, and upper left coordinates, respectively. The expression $p[i,j]=p[i\Delta\mu,j\Delta\nu](x,y,z)$ represents the pressure produced by the uniformly excited rectangular source having width $i\Delta\mu$ and height $j\Delta\nu$, where the lower left corner is located at (0,

0), and the aperture function over the rectangle $S[i,j]$ is defined by $q[i,j]=f(i\Delta\mu,j\Delta\nu)$. The total pressure $p_{\text{apod}}(x,y,z;k)$ is approximated by

$$p_{\text{apod}}(x,y,z;k) \simeq \sum_{i=1}^{N}\sum_{j=1}^{N} q[i,j](p[i,j]+p[i-1,j-1] - p[i,j-1]-p[i-1,j]). \quad (3)$$

According to Eq. (2), $p[l,0]=0$, and $p[0,l]=0$, $l=1,\ldots,N$. By utilizing these restrictions and rearranging Eq. (3),

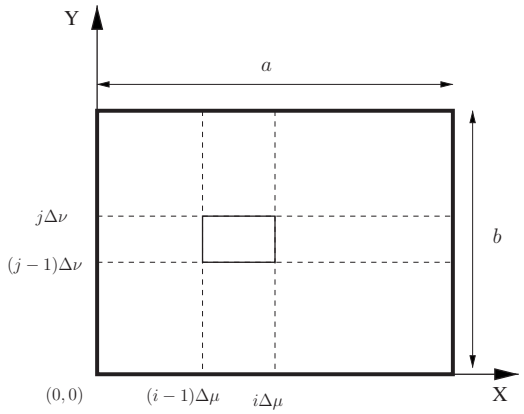


FIG. 2. The decomposition of an apodized rectangular source into smaller rectangles, where each small rectangle is $\Delta\mu$ wide and $\Delta\nu$ high. The apodization function $f(\mu,\nu)$ is defined as constant over each small rectangle.

$$p_{\text{apod}}(x,y,z;k) \simeq \sum_{i=1}^{N-1}\sum_{j=1}^{N-1} p[i,j]\left(\frac{(q[i+1,j+1]-q[i+1,j])-(q[i,j+1]-q[i,j])}{\Delta\nu\Delta\mu}\Delta\nu\Delta\mu\right)+\sum_{i=1}^{N-1} p[i,N] \times\left(-\frac{q[i+1,N]-q[i,N]}{\Delta\mu}\right)\Delta\mu+\sum_{j=1}^{N-1} p[N,j]\left(-\frac{q[N,j+1]-q[N,j]}{\Delta\nu}\right)\Delta\nu+p[N,N]q[N,N] \quad (4)$$

is obtained. Letting $N\to\infty$ such that $\Delta\mu\to 0$ and $\Delta\nu\to 0$, Eq. (4) becomes

$$p_{\text{apod}}(x,y,z;k)=p_{\text{apod1}}(x,y,z;k)+p_{\text{apod2}}(x,y,z;k)+p_{\text{apod3}}(x,y,z;k)+p_{\text{apod4}}(x,y,z;k), \quad (5)$$

where

$$p_{\text{apod1}}(x,y,z;k)=\int_0^a\int_0^b(\partial^2 f(\mu,\nu)/\partial\mu\partial\nu)p_0[\mu,\nu](x,y,z;k)d\mu d\nu,$$

$$p_{\text{apod2}}(x,y,z;k)=-\int_0^a(\partial f(\mu,b)/\partial\mu)p_0[\mu,b](x,y,z;k)d\mu,$$

$$p_{\text{apod3}}(x,y,z;k)=-\int_0^b(\partial f(a,\nu)/\partial\nu)p_0[a,\nu](x,y,z;k)d\nu,$$

and

$$p_{\text{apod4}}(x,y,z;k)=f(a,b)p_0[a,b](x,y,z;k).$$

Within these expressions, $p_0[\mu,\nu](x,y,z;k)$ is the single integral from Eq. (1), so $p_{\text{apod1}}(x,y,z;k)$ is actually a triple integral, $p_{\text{apod2}}(x,y,z;k)$ and $p_{\text{apod3}}(x,y,z;k)$ are double integrals that admit further simplification, and $p_{\text{apod4}}(x,y,z;k)$ is already simplified.

After substituting the uniformly excited FNM expression $p_0[\mu,\nu](x,y,z;k)$ from Eq. (1), $p_{\text{apod1}}(x,y,z;k)$ in Eq. (5) becomes

$$p_{\text{apod1}}(x,y,z;k)=-\frac{\rho c v_0 e^{j\omega t}}{2\pi}\sum_{i=1}^{4}\int_{n_i}^{m_i}\int_0^a\int_0^b\frac{\partial^2 f(\mu,\nu)}{\partial\mu\partial\nu}h_i \times\frac{e^{-jk\sqrt{\sigma^2+z^2+h_i^2}}-e^{-jkz}}{\sigma^2+h_i^2}d\sigma d\mu d\nu. \quad (6)$$

Since the four triple integrals in Eq. (6) all share the same structure, only the derivation of the first integral

$$p_{\text{apod1},i=1}(x,y,z;k)=-\frac{\rho c v_0 e^{j\omega t}}{2\pi}\int_{-x}^{\mu-x}\int_0^a\int_0^b\frac{\partial^2 f(\mu,\nu)}{\partial\mu\partial\nu}h_i \times\frac{e^{-jk\sqrt{\sigma^2+z^2+h_i^2}}-e^{-jkz}}{\sigma^2+h_i^2}d\sigma d\mu d\nu \quad (7)$$

is outlined here, where $p_{\text{apod1}}(x,y,z;k)=p_{\text{apod1},i=1}(x,y,z;k)+p_{\text{apod1},i=2}(x,y,z;k)+p_{\text{apod1},i=3}(x,y,z;k)+p_{\text{apod1},i=4}(x,y,z;k)$. Let $g_1(\mu)$ denote the integral

$$g_1(\mu)=\int_x^{\mu-x} y\frac{e^{-jk\sqrt{\sigma^2+z^2+y^2}}-e^{-jkz}}{\sigma^2+y^2}d\sigma, \quad (8)$$

where $h_1=y$ as indicated above. The function $g_1(\mu)$ is defined in terms of a variable in the upper limit of the integral, so the derivative of $g_1(\mu)$ with respect to μ is

$$g_1'(\mu)=y\frac{e^{-jk\sqrt{(\mu-x)^2+z^2+y^2}}-e^{-jkz}}{(\mu-x)^2+y^2}. \quad (9)$$

After integrating by parts with respect to the variable μ, $p_{\text{apod1},i=1}(x,y,z;k)$ is rewritten as

$$p_{\text{apod1},i=1}(x,y,z;k) = -\frac{\rho c v_0 e^{j\omega t}}{2\pi}\int_0^b \left[\frac{\partial f(\mu,\nu)}{\partial \nu} g_1(\mu)\Big|_0^a - \int_0^a \frac{\partial f(\mu,\nu)}{\partial \nu} g_1'(\mu)d\mu\right] d\nu. \quad (10)$$

According to Eq. (8), $g_1(0)=0$. Substituting the expression for $g_1(\mu)$ into Eq. (10) and performing an exchange of variables yield the following analytically equivalent expression for the first triple integral:

$$p_{\text{apod1},i=1}(x,y,z;k) = -\frac{\rho c v_0 e^{j\omega t}}{2\pi}\int_0^a \int_0^b \left(\frac{\partial f(a,\nu)}{\partial \nu} - \frac{\partial f(\mu,\nu)}{\partial \nu}\right) y \times \frac{e^{-jk\sqrt{(\mu-x)^2+z^2+y^2}} - e^{-jkz}}{(\mu-x)^2+y^2} d\mu d\nu. \quad (11)$$

The other three triple integrals in p_{apod1} are simplified into double integrals using the same strategy.

By substituting $p_0[\mu,b](x,y,z;k)$ and $p_0[a,\nu](x,y,z;k)$ from Eq. (1) into $p_{\text{apod2}}(x,y,z;k)$ and $p_{\text{apod3}}(x,y,z;k)$ and by using integration by parts, $p_{\text{apod2}}(x,y,z;k)$ and $p_{\text{apod3}}(x,y,z;k)$ are converted into the sum of two single integrals and two double integrals, while $p_{\text{apod4}}(x,y,z;k)$ is already a single integral. After the four terms in Eq. (5) are added and common terms are canceled, the complete FNM expression for an apodized rectangular piston is

$$p_{\text{apod}}(x,y,z;k) = -\frac{\rho c v_0 e^{j\omega t}}{2\pi}\sum_{i=1}^4 \int_0^a \int_0^b T_{1i} h_{1i} \times \frac{e^{-jk\sqrt{\alpha_{1i}^2+z^2+h_{1i}^2}} - e^{-jkz}}{\alpha_{1i}^2+h_{1i}^2} d\mu d\nu - \frac{\rho c v_0 e^{j\omega t}}{2\pi}\sum_{i=1}^4 \int_{n_i}^{m_i} T_{2i} h_{2i} \times \frac{e^{-jk\sqrt{\alpha_{2i}^2+z^2+h_{2i}^2}} - e^{-jkz}}{\alpha_{2i}^2+h_{2i}^2} d\sigma, \quad (12)$$

where the values of T_{1i} are $T_{11}=T_{12}=-\partial f(\mu,\nu)/\partial\nu$ and $T_{13}=T_{14}=-\partial f(\mu,\nu)/\partial\mu$, the values of α_{1i} are $\alpha_{11}=\alpha_{12}=\mu-x$ and $\alpha_{13}=\alpha_{14}=\nu-y$, and the values of h_{1i} are $h_{11}=y$, $h_{12}=\nu-y$, $h_{13}=x$, and $h_{14}=\mu-x$. The values of T_{2i} are $T_{21}=T_{22}=f(\sigma,b)$ and $T_{23}=T_{24}=f(a,\sigma)$, the values of α_{2i} are $\alpha_{21}=\alpha_{22}=\sigma-x$ and $\alpha_{23}=\alpha_{24}=\sigma-y$, the values of h_{2i} are $h_{21}=y$, $h_{22}=b-y$, $h_{23}=x$, and $h_{24}=a-x$, and the values of (m_i,n_i) are $(m_1,n_1)=(m_2,n_2)=(a,0)$ and $(m_3,n_3)=(m_4,n_4)=(b,0)$ for $i=1,2,3,4$.

The apodized FNM expression in Eq. (12), which contains the summation of four double integrals and four single integrals, describes the pressure generated by an apodized piston for any boundary condition. The apodized FNM expression in Eq. (12) admits further simplification if the apodization function is equal to zero on the piston edge, where

$$f(0,\nu)=0, \quad f(a,\nu)=0, \quad f(\mu,0)=0, \quad \text{and } f(\mu,b)=0. \quad (13)$$

The boundary conditions given by Eq. (13) are equivalent to setting all of the terms T_{2i} equal to 0, so the single integrals with respect to σ in Eq. (12) disappear. The resulting apodized FNM expression is the summation of the four double integrals with respect to μ and ν, so only the first sum in Eq. (12) is needed when the boundary values are all zero.

The FNM expression for the uniformly excited rectangular piston is also a special case of the apodized FNM expression in Eq. (12). For the uniform case, the apodization function is $f(\mu,\nu)=\text{rect}(\mu/a)\text{rect}(\nu/b)$ where $\text{rect}(t)=1$ if $t\in(0,1)$ and $\text{rect}(t)=0$ otherwise. The weak derivative of $f(\mu,\nu)$ is given by $\partial f(\mu,\nu)/\partial\mu=\delta(\mu)-\delta(\mu-a)$ and $\partial f(\mu,\nu)/\partial\nu=\delta(\nu)-\delta(\nu-b)$. When this apodization function is substituted into Eq. (12), all terms T_{2i} are equal to 0, and the double integrals reduce to a single integral, which is the same as $p_0[a,b](x,y,z;k)$ in Eq. (1).

B. Transient apodized FNM expression

The transient response for an apodized rectangular piston is obtained from the inverse Fourier transform[21] of Eq. (12). Defining $v(t)$ as the temporal component of the transient normal particle velocity $v(t)f(\mu,\nu)$, the transient pressure generated by an apodized rectangular piston is represented by

$$p_{\text{apod}}(x,y,z;t) = -\frac{\rho c}{2\pi}\sum_{i=1}^4 \int_0^a \int_0^b T_{1i} h_{1i} \frac{v(t-\sqrt{\alpha_{1i}^2+z^2+h_{1i}^2}/c) - v(t-z/c)}{\alpha_{1i}^2+h_{1i}^2} d\mu d\nu - \frac{\rho c}{2\pi}\sum_{i=1}^4 \int_{h_i}^{m_i} T_{2i} h_{2i} \frac{v(t-\sqrt{\alpha_{2i}^2+z^2+h_{2i}^2}/c) - v(t-z/c)}{\alpha_{2i}^2+h_{2i}^2} d\sigma, \quad (14)$$

where the values of T_{1i}, α_{1i}, h_{1i}, T_{2i}, α_{2i}, and h_{2i} are listed immediately after Eq. (12) in the previous section.

C. The Rayleigh–Sommerfeld integral

The time-harmonic pressure generated by an apodized rectangular source is also computed with the Rayleigh–Sommerfeld integral[14] via

$$p_{\text{Rayleigh}}(x,y,z;k) = \frac{j\omega\rho v_0 e^{j\omega t}}{2\pi}\int_0^a\int_0^b f(\mu,\nu)\frac{e^{-jkR}}{R}d\mu d\nu, \tag{15}$$

where $R=\sqrt{(x-\mu)^2+(y-\nu)^2+z^2}$ is the distance between the source point $(\mu,\nu,0)$ and the observation point (x,y,z). The transient pressure generated by an apodized rectangular source with a temporal excitation component $v(t)$ is given by the inverse Fourier transform of Eq. (15), which yields

$$p_{\text{Rayleigh}}(x,y,z;t) = \frac{\rho}{2\pi}\int_0^a\int_0^b f(\mu,\nu)\frac{\dot{v}(t-R/c)}{R}d\mu d\nu, \tag{16}$$

where $\dot{v}(t)$ is the time derivative of the input excitation pulse $v(t)$. The expressions for the Rayleigh–Sommerfeld integral in Eqs. (15) and (16) are analytically equivalent to the apodized FNM expressions in Eqs. (12) and (14), but the numerical properties of the two methods differ as demonstrated in the results shown below.

D. The FIELD II program

The FIELD II program[13] is a software package that computes transient and steady-state pressures generated by phased arrays and individual ultrasound transducers. The FIELD II program with the *use_rectangles* option divides each piston source into small rectangular elements and applies the far-field approximation of the spatial impulse response to each small rectangular element.[12] FIELD II specifies the apodization at the center point of each small rectangular element. The accuracy of the FIELD II program is dependent on two factors, namely, the sampling frequency and the number of small rectangular elements. With an increase in the sampling frequency or the number of small rectangular elements. FIELD II achieves smaller errors, but the computation time increases accordingly.

E. Apodization function

The apodization function selected for comparisons between the apodized FNM, the Rayleigh–Sommerfeld integral, and the FIELD II program is the product of sinusoidal functions given by

$$f(\mu,\nu) = \sin(\mu\pi/a)\sin(\nu\pi/b). \tag{17}$$

This function corresponds to the lowest order vibration mode of a rectangular membrane with fixed edges.[20,22] Equation (17) is plotted for an apodized rectangular source with each side equal to four wavelengths in Fig. 3.

The apodized FNM equations of Eqs. (12) and (14) admit additional simplification when applied to the apodization function in Eq. (17). The apodization function is the product of two sinusoidal functions, so the apodization function is separable with respect to the variables μ and ν. After the derivative of the apodization function with respect to μ and ν is substituted into Eqs. (12) and (14), the first and third double integrals contain $\int_0^b \cos(\nu\pi/b)d\nu$ and $\int_0^a \cos(\mu\pi/a)d\mu$ terms, respectively. These two integrals are exactly equal to 0, so the first and third double integrals in Eqs. (12) and (14) are equal to 0 for the apodization function in Eq. (17) at any observation point. Thus, only the second and fourth double integrals are needed. Furthermore, in Eqs. (12) and (14), the second and the fourth double integrals share several terms. Shared terms in the apodized FNM equations are always computed once and stored for use in repeated calculations.

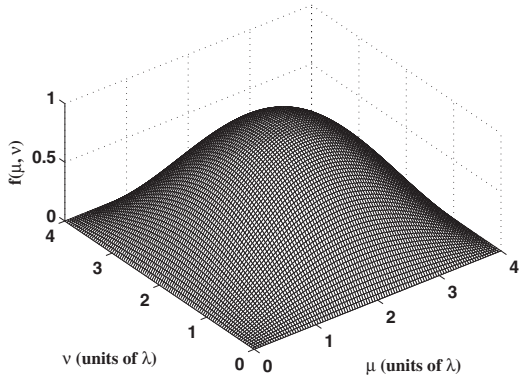


FIG. 3. The apodization function $f(\mu,\nu)=\sin(\mu\pi/a)\sin(\nu\pi/b)$ evaluated on the face of a $4\lambda\times4\lambda$ square piston. The maximum value of the apodization function is achieved when $\mu=2\lambda$ and $\nu=2\lambda$.

F. Input transient pulse

For transient calculations, the excitation pulse $v(t)$ is specified by the Hanning-weighted pulse

$$v(t) = \tfrac{1}{2}[1-\cos(2\pi t/W)]\sin(2\pi f t)\text{rect}(t/W), \tag{18}$$

where $\text{rect}(t)=1$ if $t\in(0,1)$ and $\text{rect}(t)=0$ otherwise. In the transient simulations that follow, the center frequency f_0 and the pulse duration W are $f_0=1.5$ MHz and $W=2.0$ μs, respectively, for the Hanning-weighted pulse.

G. Time space decomposition

For transient calculations, most of the computation time is expended while evaluating $v(t-\tau)$ in Eq. (14) and $\dot{v}(t-\tau)$ in Eq. (16). These terms are calculated at each time t. The variable τ is a function of the observation coordinates only, so τ represents the contribution from the spatial variable. The function $v(t-\tau)$ can be separated according to the time space decomposition approach in Ref. 21. This expression is decoupled as

$$v(t-\tau) = \text{rect}\left(\frac{t-\tau}{W}\right)\sum_{m=1}^{M} f_m(\tau)g_m(t), \tag{19}$$

where $M=6$ for transient FNM calculations with the Hanning-weighted pulse in Eq. (18), W is the length of the pulse, and the $f_m(\tau)$ and $g_m(t)$ terms are given in Table I. The transient input to Eq. (16) is decoupled in the same manner, and the corresponding $f_m(\tau)$ and $g_m(t)$ terms for $\dot{v}(t-\tau)$ are given in Table II where $M=10$ for transient Rayleigh–

TABLE I. Terms that define the time space decomposition of the Hanning-weighted pulse $v(t-\tau)$ for transient apodized FNM calculations.

Temporal basis functions $g_m(t)$
$g_1(t)=\frac{1}{2}\sin(2\pi f_0 t)$
$g_2(t)=-\frac{1}{2}\cos(2\pi f_0 t)$
$g_3(t)=-\frac{1}{2}\cos\left(\frac{2\pi t}{W}\right)\sin(2\pi f_0 t)$
$g_4(t)=\frac{1}{2}\cos\left(\frac{2\pi t}{W}\right)\cos(2\pi f_0 t)$
$g_5(t)=-\frac{1}{2}\sin\left(\frac{2\pi t}{W}\right)\sin(2\pi f_0 t)$
$g_6(t)=\frac{1}{2}\sin\left(\frac{2\pi t}{W}\right)\cos(2\pi f_0 t)$
Spatial basis functions $f_m(\tau)$
$f_1(\tau)=\cos(2\pi f_0 \tau)$
$f_2(\tau)=\sin(2\pi f_0 \tau)$
$f_3(\tau)=\cos\left(\frac{2\pi \tau}{W}\right)\cos(2\pi f_0 \tau)$
$f_4(\tau)=\cos\left(\frac{2\pi \tau}{W}\right)\sin(2\pi f_0 \tau)$
$f_5(\tau)=\sin\left(\frac{2\pi \tau}{W}\right)\cos(2\pi f_0 \tau)$
$f_6(\tau)=\sin\left(\frac{2\pi \tau}{W}\right)\sin(2\pi f_0 \tau)$

Sommerfeld calculations with the Hanning-weighted pulse. A simplified version of the time space decomposition algorithm is outlined below.

(1) Precompute and store $g_m(t)$ in advance for all values of t.
(2) Evaluate τ once for each spatial coordinate.
(3) Compute the values of each $f_m(\tau)$ term immediately after τ is calculated.
(4) Calculate the value of $v(t-\tau)$ according to Eq. (19).

By exploiting repeated calculations, this approach dramatically reduces the computation time without increasing the numerical error.

H. Error calculations

Two error metrics are used in this paper: one is the normalized error distribution $\eta(x,y,z;k)$, which describes the absolute value of the pressure difference at each spatial point for time-harmonic calculations; and the other is the normal-

TABLE II. Terms that define the time space decomposition of the derivative of a Hanning-weighted pulse $\dot{v}(t-\tau)$ for transient calculations with the apodized Rayleigh–Sommerfeld integral.

Temporal basis functions $g_m(t)$
$g_1(t)=\pi f_0\cos(2\pi f_0 t)$
$g_2(t)=\pi f_0\sin(2\pi f_0 t)$
$g_3(t)=-\pi f_0\cos\left(\frac{2\pi t}{W}\right)\cos(2\pi f_0 t)$
$g_4(t)=-\pi f_0\cos\left(\frac{2\pi t}{W}\right)\sin(2\pi f_0 t)$
$g_5(t)=-\pi f_0\sin\left(\frac{2\pi t}{W}\right)\cos(2\pi f_0 t)$
$g_6(t)=-\pi f_0\sin\left(\frac{2\pi t}{W}\right)\sin(2\pi f_0 t)$
$g_7(t)=\frac{\pi}{W}\sin\left(\frac{2\pi t}{W}\right)\sin(2\pi f_0 t)$
$g_8(t)=-\frac{\pi}{W}\sin\left(\frac{2\pi t}{W}\right)\cos(2\pi f_0 t)$
$g_9(t)=-\frac{\pi}{W}\cos\left(\frac{2\pi t}{W}\right)\sin(2\pi f_0 t)$
$g_{10}(t)=\frac{\pi}{W}\cos\left(\frac{2\pi t}{W}\right)\cos(2\pi f_0 t)$
Spatial basis functions $f_m(\tau)$
$f_1(\tau)=\cos(2\pi f_0 \tau)$
$f_2(\tau)=\sin(2\pi f_0 \tau)$
$f_3(\tau)=\cos\left(\frac{2\pi \tau}{W}\right)\cos(2\pi f_0 \tau)$
$f_4(\tau)=\cos\left(\frac{2\pi \tau}{W}\right)\sin(2\pi f_0 \tau)$
$f_5(\tau)=\sin\left(\frac{2\pi \tau}{W}\right)\cos(2\pi f_0 \tau)$
$f_6(\tau)=\sin\left(\frac{2\pi \tau}{W}\right)\sin(2\pi f_0 \tau)$
$f_7(\tau)=\cos\left(\frac{2\pi \tau}{W}\right)\cos(2\pi f_0 \tau)$
$f_8(\tau)=\cos\left(\frac{2\pi \tau}{W}\right)\sin(2\pi f_0 \tau)$
$f_9(\tau)=\sin\left(\frac{2\pi \tau}{W}\right)\cos(2\pi f_0 \tau)$
$f_{10}(\tau)=\sin\left(\frac{2\pi \tau}{W}\right)\sin(2\pi f_0 \tau)$

ized root mean square error (NRMSE), which describes the overall error performance with a single value. Let $p(x,y,z;k)$ represent the computed pressure field, and let $p_{\text{ref}}(x,y,z;k)$ represent the reference pressure field. The normalized error distribution $\eta(x,y,z;k)$ for each spatial point in time-harmonic calculations is given by

$$\eta(x,y,z;k) = |p(x,y,z;k) - p_{\text{ref}}(x,y,z;k)|/|p_{\text{ref}}(x,y,z;k)|_{\max(x,y,z)}. \quad (20)$$

The value of $\eta(x,y,z;k)$ is shown in each depiction of the error mesh. The NRMSE across all spatial points for time-harmonic calculations is

$$\text{NRMSE} = \sqrt{\sum_{x,y,z} |p(x,y,z;k) - p_{\text{ref}}(x,y,z;k)|^2} \Big/ \sqrt{\sum_{x,y,z} |p_{\text{ref}}(x,y,z;k)|^2}, \quad (21)$$

where $\Sigma_{x,y,z}$ denotes summation over all of the spatial grid points. The NRMSE for transient calculations is

$$\text{NRMSE} = \sqrt{\sum_{x,y,z,t} |p(x,y,z;t) - p_{\text{ref}}(x,y,z;t)|^2} \Big/ \sqrt{\sum_{x,y,z,t} |p_{\text{ref}}(x,y,z;t)|^2}, \quad (22)$$

where $\Sigma_{x,y,z,t}$ denotes summation over all spatial and temporal grid points. The values of the NRMSE are tabulated for each method in the Results section. The reference fields $p_{\text{ref}}(x,y,z;k)$ and $p_{\text{ref}}(x,y,z;t)$ for time-harmonic calculations and transient calculations, respectively, are calculated using the Rayleigh–Sommerfeld integral.

For near-field pressure calculations with the apodized FNM expressions and the Rayleigh–Sommerfeld integral, the numbers of abscissas in the μ and ν directions are the same for a square source. The NRMSE values for both methods are computed as the number of abscissas ranges from 2 to 100. The number of abscissas for a given NRMSE is determined by the smallest number of abscissas that has a NRMSE smaller than the desired NRMSE value. For near-field pressure calculations with the FIELD II program, the sampling frequency f_s and the number of small rectangular elements are the two factors that determine the value of the NRMSE. The sampling frequency is varied from 16 to 160 MHz with a step size of 16 MHz, and the number of small rectangular elements in each direction ranges from 10 to 60 with a step size of 5. The NRMSE for each combination of the sampling frequency and the number of small rectangular elements is then computed. The sampling frequency and the number of small rectangular elements for a given NRMSE are determined by the combination that has the smallest computation time.

I. Computational resources

The simulations are performed on a Dell Precision 360 personal computer with a 2.80 GHz Pentium 4 processor. The operating system is Windows XP professional. All simulation programs are written in the C language and run sequentially. The apodized FNM expression and the Rayleigh integral are both evaluated using Gauss quadrature.

III. RESULTS

A. Time-harmonic pressure calculations

1. Reference pressure field

The reference pressure field in Fig. 4 is computed for the apodization function shown in Fig. 3. In Fig. 4, the acoustic field is evaluated in the $x=2.0\lambda$ plane, where the grid in the $x=2.0\lambda$ plane extends from -2λ to 6λ in the y direction, and the grid spacing in the y direction is 0.1λ. The grid extends from $0.01a^2/4\lambda$ to $1.0a^2/4\lambda$ in the z direction with a spacing of $0.01a^2/4\lambda$. The computational grid in the z direction is shifted by $0.01a^2/4\lambda$ relative to the source in the $z=0$ plane for all three methods in order to avoid the most severe singularities in this location. Although the apodized FNM eliminates the worst singularities in the $z=0$ plane, the grid is nevertheless shifted slightly to reduce the problems that the Rayleigh integral and the FIELD II program encounter on the piston face. The reference pressure field shown in Fig. 4 is obtained using the Rayleigh–Sommerfeld integral evaluated with 100 000 Gauss abscissas in each direction, and the results are computed on an 81×100 spatial grid.

2. Error distributions

Figure 5 shows the normalized error distribution $\eta(x,y,z;k)$ between the reference pressure field in Fig. 4 and

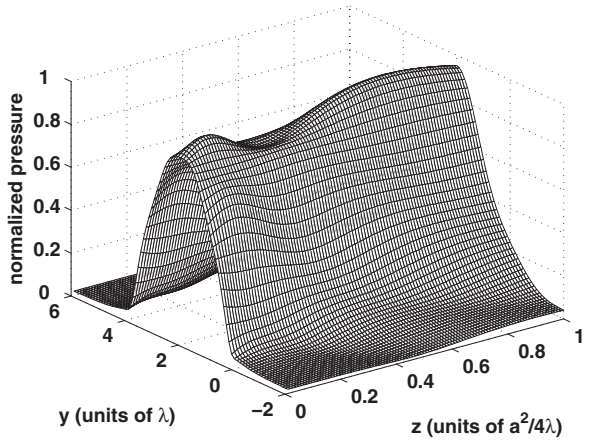
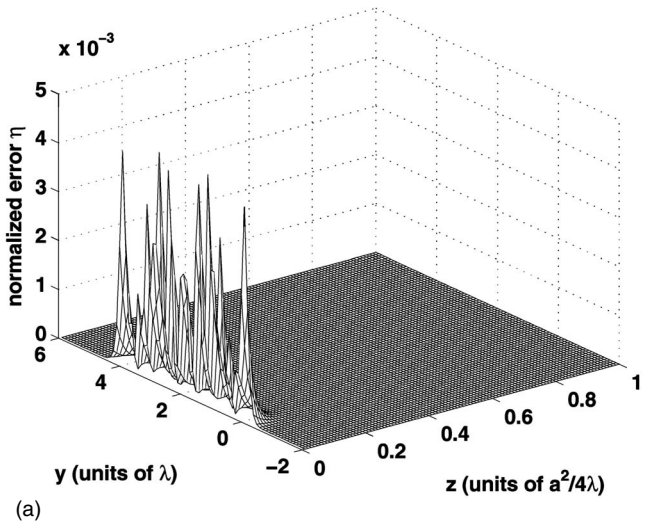
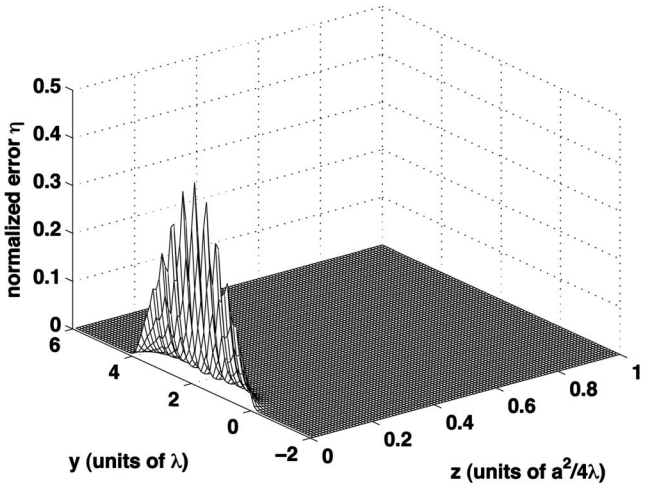


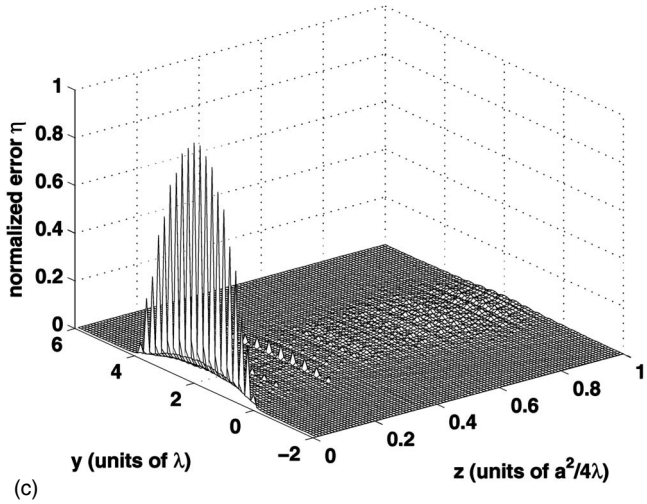
FIG. 4. Simulated reference pressure field generated by an apodized rectangular source with each side equal to four wavelengths. The results are evaluated in the $x=2.0\lambda$ plane for a time-harmonic excitation.



(a)



(b)



(c)

FIG. 5. The normalized error distribution $\eta(x,y,z;k)$ describes the difference between the reference pressure field and the computed pressure field for an apodized $4\lambda \times 4\lambda$ source. The error distribution η is plotted for (a) the apodized FNM evaluated with 16-point Gauss quadrature in each direction, (b) the apodized Rayleigh integral evaluated with 16-point Gauss quadrature in each direction, and (c) the FIELD II program evaluated with f_s=48 MHz and 30 subdivisions in each direction.

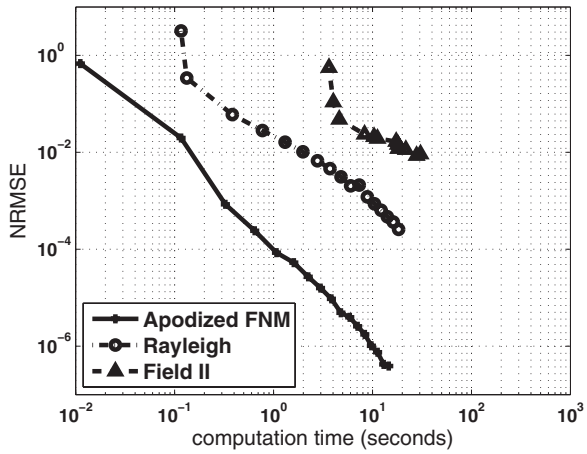


FIG. 6. Normalized root mean square error (NRMSE) plotted as a function of the computation time for time-harmonic calculations with the apodized FNM, the apodized Rayleigh–Sommerfeld integral, and the FIELD II program. This figure demonstrates that the apodized FNM achieves the smallest errors for a given computation time and that the apodized FNM requires the smallest amount of time to achieve a given error value.

the simulated pressure field computed using the apodized FNM expression, the apodized Rayleigh–Sommerfeld integral, and the FIELD II program, where each is evaluated with the apodization function specified by Eq. (17). The region near the piston face contains the largest errors for all three methods. The NRMSE for the apodized FNM evaluated with $N=16$ Gauss abscissas in each direction is 0.0005. The apodized Rayleigh–Sommerfeld integral evaluated with $N=16$ Gauss abscissas in each direction has a NRMSE of 0.0450. The FIELD II program computed with a 48 MHz sampling frequency and 30 subdivisions in each direction achieves a NRMSE of 0.0808.

3. Time versus error comparisons for a time-harmonic input

Figure 6 shows the computation times and the error values for near-field pressure calculations with a time-harmonic input applied to the apodized FNM, the Rayleigh–Sommerfeld integral, and the FIELD II program. In Fig. 6, the apodized FNM achieves the smallest errors for a given computation time, and the apodized FNM uses the least amount of time for a given error value. Comparisons between the apodized FNM, the Rayleigh integral, and the FIELD II program are summarized in Table III. To achieve a NRMSE of 0.1, the apodized FNM needs six abscissas in each direction, and the computation time is 0.07 s. The apodized FNM achieves a NRMSE of 0.01 in 0.18 s with ten abscissas in each direction. Likewise, to achieve a NRMSE of 0.1, the Rayleigh–Sommerfeld integral needs 12 abscissas in each direction, and the computation time is 0.29 s. The Rayleigh–Sommerfeld integral achieves a NRMSE of 0.01 in 2.25 s with 34 abscissas in each direction. Instead of evaluating an integral with a certain number of abscissas, the FIELD II program subdivides the aperture into small rectangular sources and calculates the impulse response at a specified sampling frequency. To achieve a NRMSE of 0.1, the FIELD II program requires a temporal sampling frequency of 64 MHz and 20 subdivisions in each direction, and the result is obtained in

TABLE III. Simulation parameters for time-harmonic calculations that achieve normalized root mean square error (NRMSE) values of 0.1 and 0.01. The parameters listed include the number of Gauss abscissas or the corresponding FIELD II parameters, the resulting computation time, and the ratio of the computation time relative to the apodized FNM for the Rayleigh integral and the FIELD II program.

	Time-harmonic calculations					
	0.1 NRMSE			0.01 NRMSE		
	Apodized FNM	Rayleigh	FIELD II	Apodized FNM	Rayleigh	FIELD II
Simulation parameters	6×6 abscissas	12×12 abscissas	$f_s=64$ MHz 20×20 subdivisions	10×10 abscissas	34×34 abscissas	$f_s=160$ MHz 50×50 subdivisions
Time (s)	0.07	0.29	4.61	0.18	2.25	27.91
Computation time relative to apodized FNM	1×	4.14×	59.43×	1×	12.50×	155.06×

4.61 s. The FIELD II program achieves a NRMSE of 0.01 with a temporal sampling frequency of 160 MHz and 50 subdivisions in each direction. This calculation is completed in 27.91 s. Thus, for a NRMSE of 0.1, the apodized FNM is 4.14 times faster than the Rayleigh integral, and the apodized FNM is 59.43 times faster than the FIELD II program. For a NRMSE of 0.01, the apodized FNM is 12.50 times faster than the Rayleigh integral and the apodized FNM is 155.06 times faster than the FIELD II program. These results are specific to the grid and piston geometry in Fig. 3 and the apodization function in Eq. (17). Although the results vary somewhat for different parameter combinations, the apodized FNM consistently achieves the best performance in time versus error comparisons. The apodized FNM is also an ideal reference for time-harmonic calculations due to the rapid convergence of Eq. (12).

B. Transient field calculations

1. Reference pressure field

The reference transient field is computed using the transient Rayleigh–Sommerfeld integral in Eq. (16), where the numerical integration is performed with 100 000 Gauss abscissas applied in both the height and width directions. The input pulse is the Hanning-weighted pulse in Eq. (18) with center frequency $f_0=1.5$ MHz and pulse length $W=2.0$ μs. The speed of sound is $c=1.5$ mm/μs, and the wavelength at the center frequency is $\lambda=1.0$ mm. The sampling frequency is 16 MHz. The transient field is evaluated in the $x=2.0\lambda$ plane, where the computational grid extends from -2λ to 6λ in the y direction and from $0.01a^2/4\lambda$ to $1.0a^2/4\lambda$ in the z direction, as shown in Fig. 1. The reference transient field is calculated on an 81×100 spatial grid evaluated at 192 time points. Figure 7 shows two transient plots of the reference pressure evaluated at times $t=1.5625$ μs and $t=3.0625$ μs.

2. Time versus error comparisons for a Hanning-weighted input pulse

Figure 8 shows the computation times and the numerical errors for the transient pressure field generated by the Hanning-weighted input pulse evaluated with the transient apodized FNM, the transient Rayleigh–Sommerfeld integral, and the FIELD II program. For both the transient apodized FNM and the transient Rayleigh–Sommerfeld integral, the simplified time space decomposition approach is applied as described in Sec. II G. In Fig. 8, the transient apodized FNM

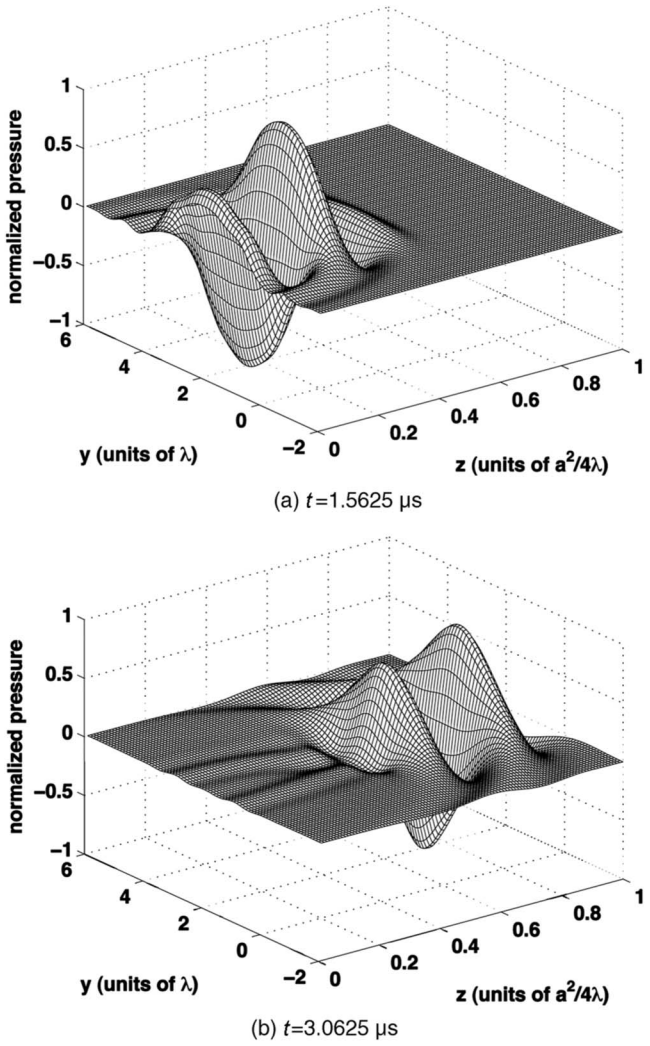


FIG. 7. Simulated reference transient field for an apodized square source excited by the Hanning-weighted pulse in Eq. (18) with $f_0=1.5$ MHz and $W=2.0\lambda$. The sides of the square source are equal to 4λ, where the wavelength is defined with respect to the center frequency f_0. The apodization function is given by Eq. (17). The transient reference pressure, evaluated in the $x=2.0\lambda$ plane, is computed with 100 000 Gauss abscissas in each direction using the Rayleigh integral. The results are plotted at (a) $t=1.5625$ μs and (b) $t=3.0625$ μs.

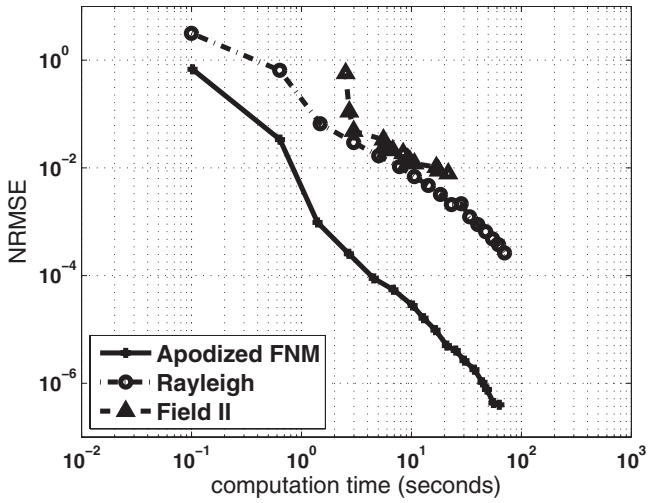


FIG. 8. Normalized root mean square error (NRMSE) plotted as a function of the computation time for transient pressure calculations evaluated with the apodized FNM, the apodized Rayleigh–Sommerfeld integral, and the FIELD II program. For the same computation time, the apodized FNM achieves the smallest errors. For the same error, the apodized FNM requires the least amount of time.

achieves the smallest errors for a given computation time, and the transient apodized FNM requires the smallest amount of time to achieve a given error. Comparisons between the transient apodized FNM, the transient Rayleigh integral, and the FIELD II program are summarized in Table IV. To achieve a NRMSE of 0.1 with the Hanning-weighted pulse, the transient apodized FNM needs eight abscissas in each direction and the computation time is 0.64 s. The transient apodized FNM achieves a NRMSE of 0.01 in 0.72 s with ten abscissas applied in each direction. To achieve a NRMSE of 0.1 with the Hanning-weighted pulse, the transient Rayleigh–Sommerfeld integral needs 14 abscissas in each direction and the computation time is 1.48 s. The transient Rayleigh–Sommerfeld integral achieves a NRMSE of 0.01 in 8.57 s with 34 abscissas in each direction. To achieve a NRMSE of 0.1 with the Hanning-weighted pulse, the FIELD II program requires a temporal sampling frequency of 64 MHz and 20 subdivisions in each source direction, and the result is obtained in 4.68 s. The FIELD II program achieves a NRMSE of 0.01 with a temporal sampling frequency of 128 MHz and 50 subdivisions in each source direction. This calculation is completed in 24.04 s. Thus, for a NRMSE of 0.1, the transient apodized FNM is 2.31 times faster than the transient Rayleigh integral and the transient apodized FNM is 4.66 times faster than the FIELD II program. For a NRMSE of 0.01, the transient apodized FNM is 11.90 times faster than the transient Rayleigh integral and the transient apodized FNM is 24.04 times faster than the FIELD II program. Once again, these results are specific to the grid, apodization function, and piston geometry utilized in this paper. Overall, the transient apodized FNM achieves the best performance in time versus error comparisons with the Hanning-weighted pulse.

IV. DISCUSSION

A. Large-scale computation

Despite the dramatic increase in processor speeds and computer memory in recent decades, near-field pressure calculations that numerically evaluate the Rayleigh–Sommerfeld integral are still very slow. This is especially evident in calculations performed on large grids, which take an hour or more on the fastest modern desktop computers. For example, more than 1 h was needed in a recent paper to compute the reference pressure field in the Fresnel zone for a single element using the Rayleigh–Sommerfeld approach.[20] More recently, near-field pressure calculations with the Rayleigh–Sommerfeld approach for a single element required 117 h per simulation to achieve a NRMSE of 10^{-7}. Thus, even for simulations of single elements performed in the near-field region, traditional computational methods perform poorly, and faster methods such as the FNM are needed.

Computation times with the Rayleigh–Sommerfeld integral can also be excessively long for phased array simulations. Extending the results shown in Tables III and IV and Figs. 6 and 8 to a linear phased array with 1024 rectangular elements evaluated on a computational grid with $100\times100\times100$ points in the x, y, and z directions, the Rayleigh–Sommerfeld integral will require about 10 h for a 0.1 NRMSE and about 80 h for a 0.01 NRMSE. For the same array evaluated on the same grid, the FNM will require only 2.5 h for a 0.1 NRMSE and 6.5 h for a 0.01 NRMSE. Thus, for large-scale computations, the time savings achieved by

TABLE IV. Simulation parameters for transient calculations that achieve normalized root mean square error (NRMSE) values of 0.1 and 0.01. The parameters listed include the number of Gauss abscissas or the corresponding FIELD II parameters, the resulting computation time, and the ratio of the computation time relative to the apodized FNM for the Rayleigh integral and the FIELD II program.

	Transient calculations					
	0.1 NRMSE			0.01 NRMSE		
	Apodized FNM	Rayleigh	FIELD II	Apodized FNM	Rayleigh	FIELD II
Simulation parameters	8×8 abscissas	14×14 abscissas	f_s=64 MHz 20×20 subdivisions	10×10 abscissas	34×34 abscissas	f_s=128 MHz 50×50 subdivisions
Time (s)	0.64	1.48	2.98	0.72	8.57	17.26
Computation time relative to apodized FNM	1×	2.31×	4.66×	1×	11.90×	24.04×

the apodized FNM is substantial. The apodized FNM is also an ideal reference for both time-harmonic and transient calculations.

B. Time and error comparisons

In addition to the results presented in Tables III and IV and Figs. 6 and 8 that compare the Rayleigh–Sommerfeld integral and the apodized FNM, the individual computation times and error values were also determined as a function of the number of abscissas. Those results show that, for a given number of abscissas, the apodized FNM achieves much smaller errors than the Rayleigh–Sommerfeld integral, and the computation time of the apodized FNM is slightly less than that of the Rayleigh integral for the same number of abscissas. Therefore, the rapid convergence of the apodized FNM is responsible for the enhancement in performance relative to other methods. However, fair comparisons of these methods require the evaluation of both times and errors, as shown in Tables III and IV and Figs. 6 and 8. Otherwise, a method that generates large errors in a similar amount of time or a method that requires much more time to achieve a similar error could be incorrectly identified as having comparable performance. Thus, the method with superior performance should consistently achieve the smallest error in the shortest time.

C. Apodization functions

Equations (12) and (14) can be applied to a large class of apodization functions. The only requirement for the apodization functions is that $\partial f(\mu,\nu)/\partial\mu$ and $\partial f(\mu,\nu)/\partial\nu$ exist. Other approaches have different requirements. For example, Fresnel approximations have been obtained for several apodization functions including exponential, sinusoidal, Gaussian, or error-function apodization functions.[20] One restriction for these Fresnel approximations is that the surface velocity distribution should be separable such that $f(\mu,\nu)=f_x(\mu)f_y(\nu)$. The advantages of the apodized FNM expressions in Eqs. (12) and (14) are that (1) separability of the apodization function is not required and (2) accurate results are obtained very close to the piston face.

Pressures generated by a Gaussian apodization function,[23–25] given by $f(\mu,\nu)=e^{-(u-u_0)^2/2\sigma_\mu^2}e^{-(\nu-\nu_0)^2/2\sigma_\nu^2}$ where μ_0, ν_0, σ_μ, and σ_ν are constants, were also evaluated with Eq. (12). When the Gaussian apodization function is applied, the entire expressions in Eqs. (12) and (14) are needed because the T_{2i} terms are nonzero in this case. Furthermore, the integration limits of the single integrals are either $(0,a)$ or $(0,b)$, so the abscissas and weights computed for the single integrals can also be used for the double integrals. The computation time needed for all terms in Eqs. (12) and (14) is only slightly more than that needed to evaluate only the double integrals, especially when the number of abscissas is larger than ten in one direction.

D. Time space decomposition

The expressions for the transient near-field pressures are computed with the simplified time space decomposition approach outlined in Sec. II G for both the transient apodized FNM expression and the transient Rayleigh–Sommerfeld integral in Eqs. (14) and (16), respectively. The simplified time space decomposition approach provides an important advantage in that, unlike the original time space decomposition method described by Kelly and McGough,[18] both one-dimensional (1D) and 2D near-field diffraction integrals are readily computed. The original time space decomposition method of Kelly and McGough[18] stores a three-dimensional (3D) K matrix that captures the arrival and departure information for a finite duration pulse, where the transient pressure propagation is described by a single integral FNM expression. Extending the original time space decomposition to handle transient calculations with the 2D apodized FNM expressions in Eq. (14) or the Rayleigh–Sommerfeld integral in Eq. (16) involves replacing the 3D K matrix with a complicated five-dimensional K matrix. With the simplified time space decomposition approach, no matrix storage is needed. Thus, the algorithm complexity and the memory usage are greatly reduced. The simplified time space decomposition in Sec. II G is preferred for transient calculations with 2D integrals, where a significant reduction in the computation time is achieved relative to other approaches without increasing the numerical error. For transient calculations with 1D integrals, there is a trade-off between the simplified time space decomposition and the original time space decomposition method. The original time space decomposition requires more memory, is more complicated to program, and is slightly faster, while the simplified time space decomposition is much easier to program.

V. CONCLUSION

A 2D FNM for calculating the near-field pressure generated by an apodized rectangular piston is derived and evaluated, and a simplified time space decomposition technique is outlined for transient apodized FNM and transient apodized Rayleigh–Sommerfeld calculations. Comparisons between the apodized FNM, the Rayleigh–Sommerfeld integral, and the FIELD II program show that the apodized FNM achieves the smallest error in the shortest time for both steady-state and transient calculations. For time-harmonic calculations with a NRMSE of 0.01, the apodized FNM is 4.14 times faster than the Rayleigh–Sommerfeld integral and 59.43 times faster than the FIELD II program. For a NRMSE of 0.01, the time-harmonic apodized FNM is 12.50 times faster than the Rayleigh–Sommerfeld integral and 155.06 times faster than the FIELD II program. For transient calculations with a NRMSE of 0.1, the apodized FNM is 2.31 times faster than the Rayleigh–Sommerfeld integral and 4.66 times faster than the FIELD II program. For a NRMSE of 0.01, the transient apodized FNM is 11.90 times faster than the Rayleigh–Sommerfeld integral and 24.04 times faster than the FIELD II program. In addition, the apodized FNM converges much faster than the Rayleigh–Sommerfeld integral and the FIELD II program. Thus, the apodized FNM is ideal for fast near field pressure calculations and also for accurate reference calculations.

ACKNOWLEDGMENTS

The authors would like to thank James F. Kelly and Xiaozheng Zeng for thoughtful discussions. This work was supported in part by NIH R01 CA093669, NIH R21 CA121235, NSF 0634786, and NIH P01 CA042745.

[1]G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[2]M. Greenspan, "Piston radiator: Some extensions of the theory," J. Acoust. Soc. Am. **65**, 608–621 (1979).
[3]C. Lafon, F. Prat, J. Y. Chapelon, F. Gorry, J. Margonari, Y. Theillere, and D. Cathignol, "Cylindrical thermal coagulation necrosis using an interstitial applicator with a plane ultrasonic transducer: *In vitro* and *in vivo* experiments versus computer simulations," Int. J. Hyperthermia **16**, 508–522 (2000).
[4]D. R. Daum and K. Hynynen, "A 256-element ultrasonic phased array system for the treatment of large volumes of deep seated tissues," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1254–1268 (1999).
[5]Y. Li and J. A. Zagzebski, "A frequency domain model for generating B-mode images with array transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 690–699 (1999).
[6]G. Cincotti, G. Cardone, P. Gori, and M. Pappalardo, "Efficient transmit beamforming in pulse-echo ultrasonic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1450–1458 (1999).
[7]Y. Lin, J. M. Dodson, J. D. Hamilton, J.-U. A. Kluiwstra, C. Cain, and K. Grosh, "Theory and experiment for the design of piezoeletric element for phased arrays," Proc.-IEEE Ultrason. Symp. **2**, 1697–1700 (1997).
[8]J. L. Prego Borges, F. Montero de Espinosa, J. Salazar, J. Garcia-Alvarez, J. A. Chavez, A. Turo, and M. J. Garcia-Hernandez, "Diffraction aperture non-ideal behaviour of air coupled transducers array elements designed for NDT," Ultrasonics **44**, 667(E)–672(E) (2006).
[9]P. R. Stepanishen, "Transient radiation from pistons in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1629–1638 (1971).
[10]F. Oberhettinger, "On transient solutions of the 'baffled piston' problem," J. Res. Natl. Bur. Stand., Sect. B **65B**, 1–6 (1961).
[11]J. C. Lockwood and J. G. Willette, "High-speed method for computing the exact solution for the pressure variations in the nearfield of a baffled piston," J. Acoust. Soc. Am. **53**, 735–741 (1973).
[12]J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[13]J. A. Jensen, "Field: A program for simulating ultrasound systems," Med. Biol. Eng. Comput. **34**, 351–353 (1996).
[14]K. B. Ocheltree and L. A. Frizzell, "Sound field calculation for rectangular sources," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 242–248 (1989).
[15]R. J. McGough, "Rapid calculations of time-harmonic nearfield pressures produced by rectangular pistons," J. Acoust. Soc. Am. **115**, 1934–1941 (2004).
[16]D. Chen, J. F. Kelly, and R. J. McGough, "A fast nearfield method for calculations of time-harmonic and transient pressures produced by triangular pistons," J. Acoust. Soc. Am. **120**, 2450–2459 (2006).
[17]R. J. McGough, T. V. Samulski, and J. F. Kelly, "An efficient grid sectoring method for calculations of the nearfield pressure generated by a circular piston," J. Acoust. Soc. Am. **115**, 1942–1954 (2004).
[18]J. F. Kelly and R. J. McGough, "A fast time-domain method for calculating the nearfield pressure generated by a pulsed circular piston," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1150–1159 (2006).
[19]X. Zeng and R. J. McGough, "Evaluation of the angular spectrum approach for simulations of nearfield pressures," J. Acoust. Soc. Am. **123**, 68–76 (2008).
[20]T. D. Mast, "Fresnel approximations for acoustic fields of rectangularly symmetric sources," J. Acoust. Soc. Am. **121**, 3311–3322 (2007).
[21]J. F. Kelly and R. J. McGough, "An annular superposition integral for axisymmetric radiators," J. Acoust. Soc. Am. **121**, 759–765 (2007).
[22]T. L. Szabo, "Generalized fourier transform diffraction theory for parabolically anisotropic media," J. Acoust. Soc. Am. **63**, 28–34 (1978).
[23]J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).
[24]D. Ding, Y. Zhang, and J. Liu, "Some extensions of the Gauss beam expansion: Radiation fields of the rectangular and the elliptical transducer," J. Acoust. Soc. Am. **113**, 3043–3048 (2003).
[25]B. D. Cook and W. J. Arnoult III, "Gaussian-Laguerre/Hermite formula for the nearfield of an ultrasonic transducer," J. Acoust. Soc. Am. **59**, 9–11 (1976).

Subjective and objective evaluations of a scattered sound field in a scale model opera house

Jong Kwan Ryu and Jin Yong Jeon[a)]
School of Architectural Engineering, Hanyang University, Seoul 133-791, Korea

(Received 7 March 2007; revised 11 June 2008; accepted 11 June 2008)

Scattered sound fields in an opera house were objectively and subjectively evaluated through acoustical measurements in a 1:10 scale model and through auditory preference tests. Acoustical characteristics were measured in the stalls area with and without diffusers, both on the sidewalls close to the proscenium and in the soffit of the side balcony. Installed diffusers reduced the initial time delay gap and amplitude of the first reflected sound, and decreased sound pressure level (SPL), reverberation time (RT), and early decay time (EDT) at most seats due to the increased scattering and absorption. After diffuser installation, C_{80} and 1-IACC$_{E3}$ increased at the front seats and decreased at the rear seats. Subjective evaluations showed that the preference of scattered sound fields correlates highly with loudness and reverberance. It was also found that EDT and SPL are dominant parameters describing subjective preference for scattered sounds in this experimental condition. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2956474]

PACS number(s): 43.55.Br, 43.55.Dt, 43.55.Hy, 43.55.Fw [NX] Pages: 1538–1549

I. INTRODUCTION

Sound diffusion has a major effect on a hall's acoustical quality. The diffusive surface(s) in a hall can prevent "acoustic glare"[1] and tone coloration[2] by strong reflection and flutter echo from multiple reflections. The effects of sound diffusion on room acoustics have been investigated for over 30 years.[3–6] Most studies have been related to the design and measurement of the sound diffusing surfaces and the prediction and quantification of scattering and diffusion coefficients. Other studies have focused on the effects of scattering on geometric room-acoustic modeling in order to enhance prediction accuracy.[7–9]

Recent attempts to quantify scattering and diffuse reflections caused by diffuse surfaces have led to the establishment of international standards for the measurement of the scattering coefficient.[10–13] Some results have indicated that acoustical factors such as initial time delay gap (ITDG) and interaural cross correlation (IACC) vary in scattered sound fields affected by the array of circular columns as diffusers.[14] Chiles and Barron's investigation[15] showed that diffusers slightly reduce the scatter of reverberation time and that the decay curve with diffusers is more linear with less deviation from the best-fit line.

Studies on the perception of diffusion have also been conducted. Ando[16] found that subjective diffuseness decreases as IACC increases, being independent of the frequencies between 250 Hz and 4 kHz. Torres *et al.*[17] investigated the audibility of diffusion in room-acoustics auralization. They found that changes in the diffusion coefficient are clearly audible within a wide frequency region. Takahashi and Takahashi[18] investigated the subjective effects of scattering periodic-type diffusers. They found that as the range of effective diffusivity in a frequency scale increases, the ratio of the perception of the difference between specular and diffuse reflections increases. They also found that as the listening position becomes closer to the diffuser, subjective tonal effects of the responses increase.

Up to this point, two parameters have been used as objective indices for indicating the degree of diffusion in sound fields. One is the sound diffusivity index with visual inspection.[19] The other is 1-IACC$_{L3}$, which was proposed by Hidaka *et al.*[20] and is related to listener envelopment. However, these objective indices were not fully verified through measurement in sound fields, and little work has been done to quantify the degree of diffusion in real sound fields. Therefore, there is a need for research on subjective and objective diffusion in real sound fields.

In this study, the effect of diffusers in a 1:10 scale model opera house was investigated both subjectively and objectively. The purpose of this study was to evaluate how installing diffusers on the sidewalls close to the proscenium arch, the sidewalls of stalls, and balcony fronts affects acoustical parameters and impulse responses, and to evaluate the quality of scattered sound in terms of subjective preference judgments.

II. SCALE MODEL MEASUREMENTS

A. Measurement setup

A 1:10 scale model of the Gimhae Arts Hall (GAH), a multipurpose hall, was built for acoustical measurements. Table I gives the architectural details of GAH. The scale model consists of the following structural elements: walls, floor, ceiling, seats, audience, and absorption banners. The materials of each element were selected through absorption coefficient measurements in a 1:10 reverberation chamber according to ISO 354.[21]

[a)]Author to whom correspondence should be addressed. Electronic mail: jyjeon@hanyang.ac.kr.

TABLE I. GAH architectural details.

Use	Concert, opera, drama, and speech events
Plan type	Shoebox+horseshoe
Seats	Total 1484 seats first and second (stalls): 864, third: 313, fourth: 307
Dimension	$W \times L \times H$: $28.5 \times 31.3 \times 20.4$ m^3
Volume	15,325 m^3 (main hall and orchestra enclosure)
Proscenium	W: 18.0 m, H: 12.0 m
Pit	Depth: 7.7 m, height: 3.0 m, area: 169 m^2
Adjustable elements	Absorption banner, orchestra enclosure, movable orchestra pit

Hemisphere and polygon diffusers made with lacquered wood were used to provide scattering surfaces in this study. The height of the hemisphere diffusers ranged from 15 to 25 mm and the height of the polygon diffusers was 30 mm. Jeon *et al.*[11] found that the scattering coefficient of a 17.5 mm (17.5 cm for an equivalent full-size hall) high, hemisphere-shaped diffuser is at its maximum when the coverage density reaches 50%. The coverage density of diffusers in this study was determined to be about 40%. The scattering coefficients of the hemisphere and polygon diffusers were measured in a 1:10 scale reverberation chamber with two tweeter speakers and three 1/8 in. microphones according to ISO 14791-1.[12] During one rotation of the turntable, 72 coherent averages (n) of the impulse responses were measured. Six reverberation times were measured at each of the three microphone locations, resulting in 18 reverberation times that were averaged. An MLSSA 2000 system was used for the reverberation time measurements.

As shown in Fig. 1, for frequencies from 500 Hz to 3.15 kHz, the averaged scattering coefficients of the hemisphere and polygon diffusers are approximately 0.6 and 0.9, respectively. The higher scattering coefficients of the polygon diffuser may be due to edge effects. Edge effects can occur due to variations in the height of the sample along the edge of the test sample, and can lead to scattering coefficients greater than 1.[12]

The binaural impulse responses were measured in the stalls both with and without diffusers on the walls near the proscenium and on the soffit of the side balcony. The installed hemisphere diffusers were 15, 17.5, 20, 22.5, and 25 mm high. The coverage density of the diffusers was about 40%. In addition, polygon diffusers that were 30 mm high were installed only on the sidewalls in the stalls. In this study, the heights of the diffusers ranging from 15 to 30 cm for full-size equivalent dimension were nearly equal to the wavelength of 1–2 kHz sounds. The scattering coefficient of the both shaped diffusers was highest in the 1–2 kHz octave band (from 800 to 2200 Hz). Therefore, the effective frequency range of the both shaped diffusers was assumed to be 1–2 kHz in terms of scattering performance. Figure 2 shows the scale model opera house with and without both shaped diffusers. It was assumed that the hall had a "specular sound field" without diffusers and a "scattered sound field" with diffusers.

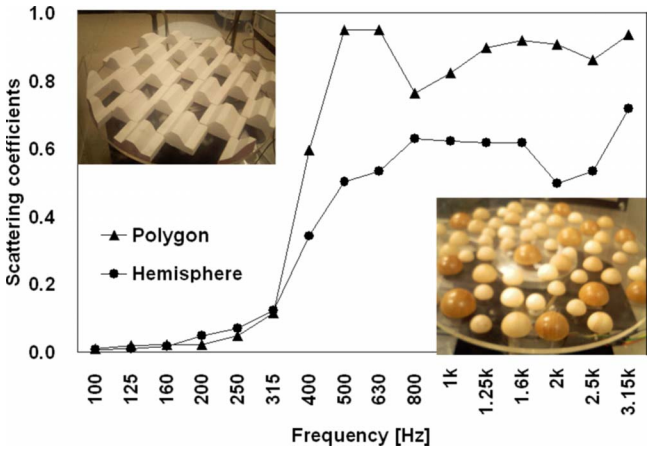
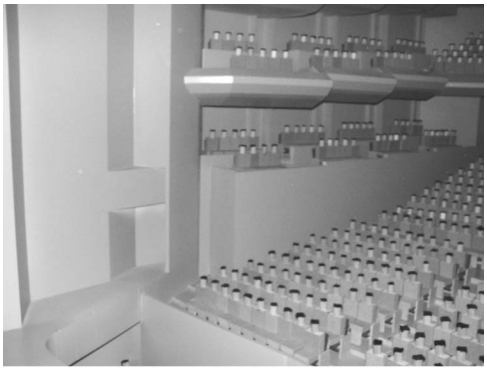
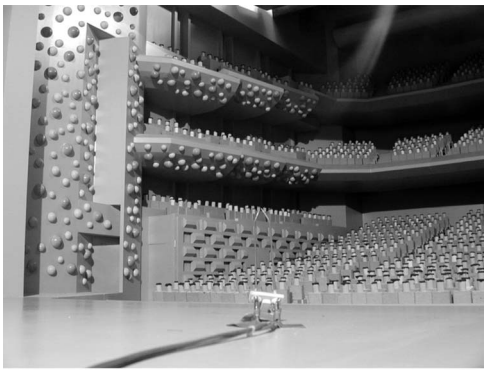


FIG. 1. (Color online) Scattering coefficients of diffusers determined from measurements according to ISO 17497-1. The figure shows average values obtained from 500 Hz to 3.15 kHz.



(a)



(b)

FIG. 2. Diffuser arrangement in the scale model: (a) no diffusers; (b) diffusers on the walls adjacent to the orchestra pit, on the sidewalls, and on the soffit of the side balcony.

An electric-spark source, and a scale model head and torso were used as the source and receiver, respectively. Xiang and Blauert[22,23] used the binaural, artificial heads with precisely modeled pinna and two tiny microphones inside the ear canals in 1:10 room-acoustic scale modeling. They also proposed air absorption compensation by replacing air by nitrogen and an advanced binaural measurement technique such as using m-sequences.

In the present study, air absorption of sound impulses was not considered in the scale model measurement because this investigation focused on the sound fields in the hall with and without a diffuser on the walls. The electric-spark source in the present study covers a high frequency range up to

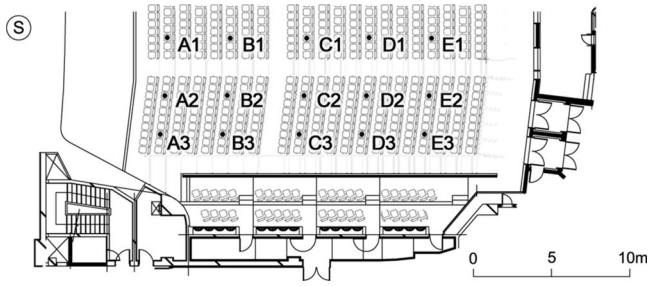


FIG. 3. Source and receiver locations.

around 80 kHz and has satisfactory omnidirectional characteristics. When the sound pressure level of impulse responses from the sparks source was measured at 10 cm (1 m full size at a scale of 1:10) and every 30° from the sources, the standard deviation of the sound pressure level was less than 1 dB in the 1–2 kHz octave band of the full size. The scale model head and torso consisted of a 21 mm diameter, styrofoam head, and two 1/8 in. microphones with a nose cone installed at each ear location for binaural measurement. The signals imported to both actual dummy and 1:10 scale model heads were measured in a semianechoic chamber. Pink noise and the music signal used in the subjective test were recorded from directional loudspeakers at every 18°, from 0° to 180°, in horizontal plane. The distances between the heads and loudspeakers were 1.5 and 0.15 m for the actual and scale models, respectively. Signals for the scale model head were scaled up (duration was shortened), and after recording it was transformed into an actual scale. Analysis of IACC by integrating whole duration of signals was made at 1, 2, and 4 kHz octave bands for the interaural time delay between −1 and +1 ms. Results showed that the difference in IACC between both heads were 0.03 (r=0.99, $p<0.01$), 0.08 (r =0.85, $p<0.01$), and 0.08 (r=0.86, $p<0.01$) at 1, 2, and 4 kHz octave bands, respectively, which are average values from all the incidence degrees for the signals.

Impulse responses generated by the spark sources were recorded using an amplifier, an AD/DA converter, and a laptop computer. The measurement system used in this study covered high frequency ranges up to 98 kHz by measuring impulse response at a 192 kHz sampling rate. The signal to noise ratio of the impulse response from the system used in this study was about 40 dB in the middle frequency range of the full size. The sampling rate of 192 kHz to cover up to 8 kHz of the full size is suitable for the evaluation of binaural parameters, other relevant room-acoustic parameters up to this frequency, and auralization by selecting frequency-limited source signals for convolution.[24]

The impulse responses, averaged from five trials at each seat, were used for calculating acoustical parameters and for convoluting music sources of the subjective tests. The source and receiver positions are shown in Fig. 3. The stage source was placed at a distance of 3 m from the front edge of the stage and 1 m from the centerline. The height of the sources was 1.5 m above the floor. The 15 receivers were located in the stalls, positioned 1.2 m above the floor. Measurements were made in an occupied condition using a model audience. The stage house was separated from the side and back stages by velour curtains in order to eliminate the effect of a long reverberation time, which would dominate the decay of sound within the auditorium. The ceiling of the stage house was also covered with velour. The orchestra pit was occupied with 50 musician models. To avoid confusion, in the later results and discussion, dimensions and frequencies were given as the full-size equivalents to an actual hall unless otherwise indicated.

B. Acoustical parameters

The values of the following acoustical parameters were determined with and without diffusers on the sidewalls: sound pressure level (SPL), reverberation time (RT), early decay time (EDT), clarity (C_{80}), center time (T_s), ITDG, and IACC. In order to consider the early and late parts of scattered reflection, 80 ms was used as the integration limit that divides the early and late parts of the impulse response. A discussion about the integration limit is described in Appendix A. Measurement of lateral energy was not made because there was no suitable device for measuring it in the 1:10 scale model. All parameter values except for IACC are given as values derived by averaging the results of 1 and 2 kHz octave bands according to the effective frequencies of the installed diffusers. IACC was measured in three octave bands with center frequencies at 0.5, 1, and 2 kHz. IACC values with different integration periods were also used: 1-IACC$_{E3}$ (0–80 ms) and 1-IACC$_{L3}$ (80–1000 ms).[25] It is important to note that the correlation coefficient between IACC$_{E3}$ and IACC$_{E2}$ (0.5 and 1 kHz), and IACC$_{L3}$ and IACC$_{L2}$ (0.5 and 1 kHz) were 0.98 ($p<0.01$) and 0.90 ($p<0.01$), respectively.

C. Results

Figure 4 shows the differences between the scattered and specular sound fields for each acoustical parameter as a function of source-receiver distance. The differences were determined by subtracting the specular values from the scattered values. As shown in Fig. 4, when the diffusers were installed, SPL decreased by about an average of 1 dB at most seats, and decreased by a larger amount at rear receiver positions and near the sidewall. RT also decreased by about an average of 0.04 s at most seats. EDT decreased at the front seats by an average of approximately 0.07 s. C_{80} increased by an average of 0.5 dB at the front seats (A and B rows), and decreased at the rear seats near the installed diffusers. Both the 1-IACC$_{E3}$ and 1-IACC$_{L3}$ values increased at the front and side seats near the source. The difference of 1-IACC$_{E3}$ with and without diffusers was greater than that of 1-IACC$_{L3}$.

Figure 5 shows the ITDG and amplitude of the first reflection at each receiver position in both scattered and specular sound fields. As shown in Fig. 5, ITDG was shortened by the presence of the diffusers due to the proximity of the reflecting surfaces to the receiver positions. The diffusers were mounted on the flat walls resulting in the reflecting surfaces being a little closer to the receiver positions, and in

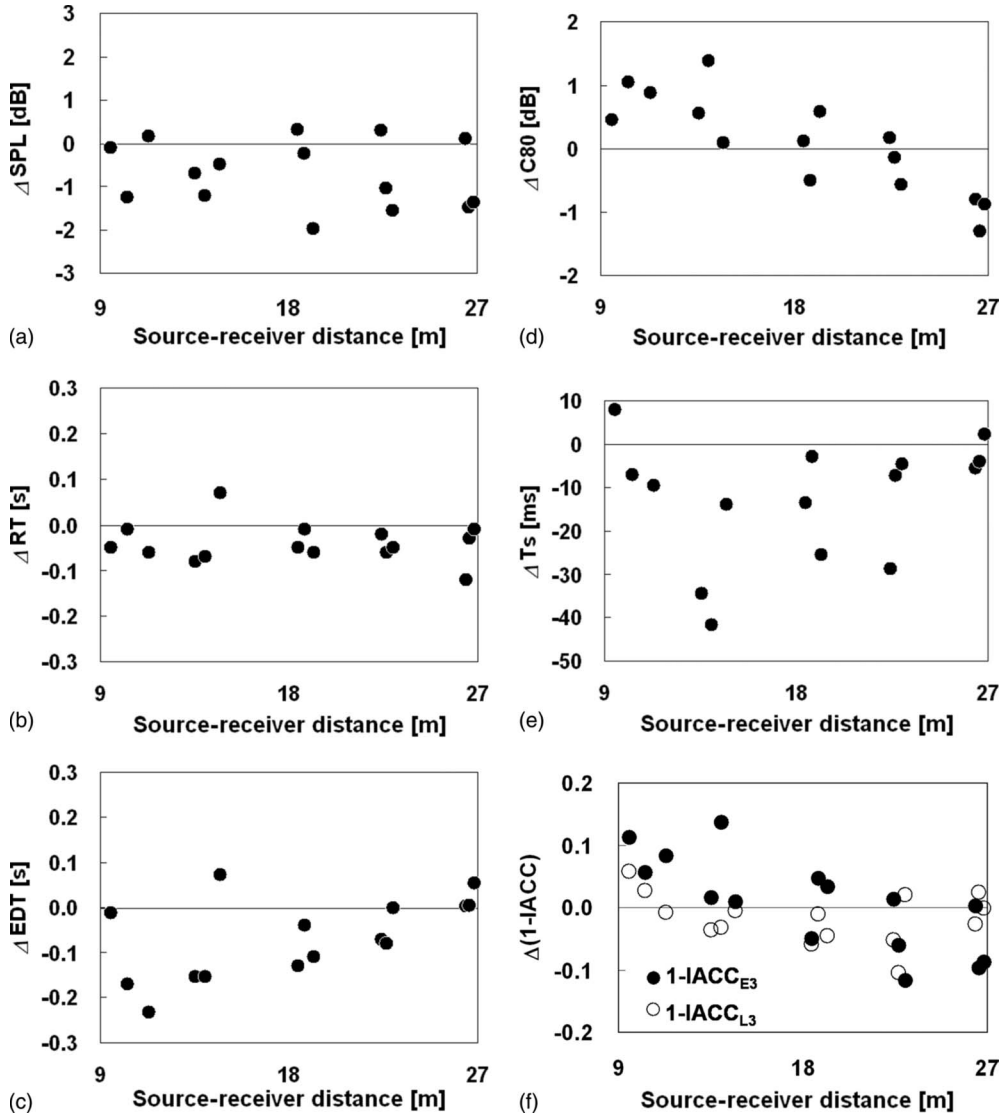


FIG. 4. Differences in several acoustical parameters (averaged over 1 and 2 kHz) between scattered and specular sound fields for various source-receiver distances (Δ in the vertical axis was calculated by subtracting the "specular" values from the "scattered" values).

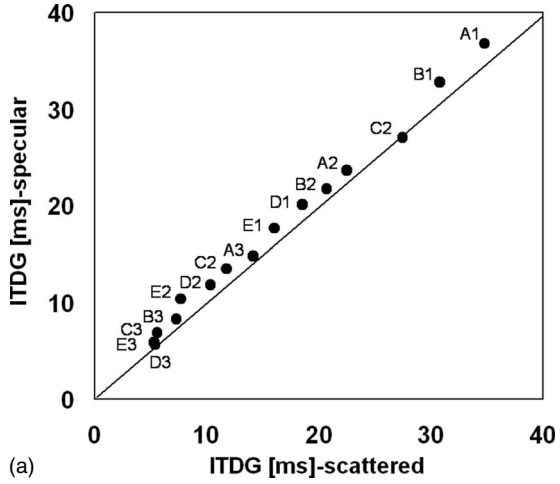
the paths of the reflections being shorter than those without the diffusers. The first reflection follows the shortest possible path according to Femat's principle.

It was also found that the SPL of the first reflection of scattered sound was smaller than that of the specular sound at most receiver positions. The amplitude of the first reflection decreased at the side and rear positions (B3, C3, row D, and row E) near the installed diffusers.

Figure 6 shows the relationship between the ITDG and amplitude of the first reflection. As shown in Fig. 6, the amplitude of the first reflection decreased as ITDG increased in both the scattered and specular sound fields. The effect of the diffusers on the amplitude of the first reflection was mainly observed at the side and rear positions where the ITDG was shorter than 20 ms.

Table II shows the average and standard deviations of the sound pressure levels of the impulse responses for various integration intervals and reverberation times at 15 seat positions. In Table II, the SPL for each integration interval was calculated with a rms value. In all cases, the results showed that the average sound pressure levels of the scattered sound field were about 1 dB lower than that of the specular sound field. The standard deviation of the scattered sound varied according to the integration time. For early reflections (from 0 to 80 ms), the standard deviation of the scattered sound level was larger than that of the specular sound level, while the standard deviation of the late part (from 80 to ∞ ms) was larger for the specular sound level.

Figure 7 shows the differences in the sound energy levels between the scattered and specular sound fields for various positions. Here, the sound energy is divided into early (3–80 ms) and late (80–∞ ms) parts. The direct sound energy (0–3 ms) was excluded in the calculation in order to eliminate its effect on the total sound energy. The results show that differences (scattered—specular) in the early sound were larger than late sound at the front seats (seats A1-3 and B1-3). This is expressed as $\Delta E_{\text{early}} > \Delta E_{\text{late}}$. However, the opposite was observed in the middle and rear seats (seats C1-3, D1-3, and E1-3), and is expressed as $\Delta E_{\text{early}} < \Delta E_{\text{late}}$. This explains why ΔC_{80} decreased with increasing source-receiver distance (see Fig. 4).



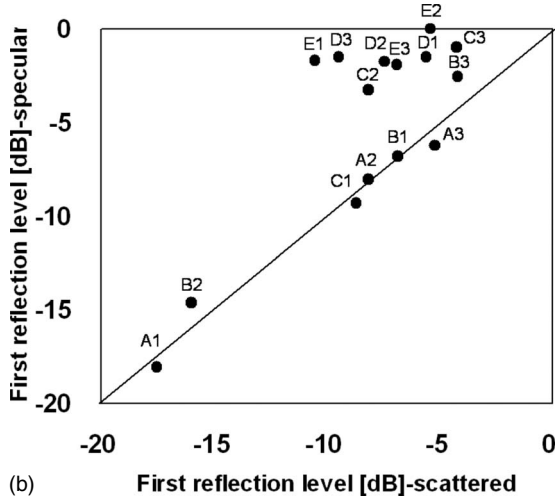


FIG. 5. Delay time (top) and sound level (bottom) of the first reflection for scattered and specular sound fields.

Figure 8 shows the backward integration curves of the scattered and specular sound fields at position A2. As shown at the other receiving positions, the scattered sound has a smoother decay curve than specular sound in the early part of reflection. A similar result was shown in a previous study,[15] where the decay curve with diffusers was more linear, with a smaller deviation from the best-fit line in early and late parts. This is due to the diffuser replacing the specular reflection with scattered sounds. However, there was little difference in the latter part of the decay curve between the scattered and the specular sound. These results show that the diffusers did not have a significant effect on the decay curves.

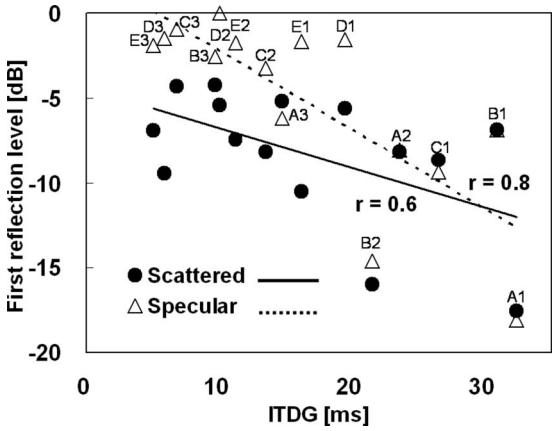


FIG. 6. Relationship between ITDG and amplitude of the first reflection.

III. SUBJECTIVE TESTS

A. Experiment 1

1. Method

An auditory experiment was conducted to investigate the subjective preference of sound fields with and without diffusers installed on the sidewalls. The scattered and specular sound fields for each seat were compared using a two-alternative forced choice method. O-X and X-O pairs (O, with diffuser; X, without diffuser) were prepared at all 15 seats totaling 30 sound-source pairs. Subjects responded to five questions pertaining to overall impression (preference), loudness, clarity, reverberance, envelopment, and apparent source width.[26] The subjects included 20 senior university music students and 11 members of a classical music club. They were chosen because they were assumed likely to have experience making judgments about the acoustical quality of sound fields. The binaural impulse responses recorded in the 1:10 scale model hall were transformed into real scale by a resampling method. The impulse responses were then convolved with a music source recorded in an anechoic cham-

TABLE II. Average and standard deviation of sound pressure levels of impulse responses for various integration intervals and reverberation times at 15 seat positions.

		SPL (dB)			
		0–80 ms	80–∞ ms	0–∞ ms	RT (s)
Scattered sound	Average	92.8	76.3	83.2	0.94
	Standard deviation	1.7	1.6	1.6	0.08
Specular sound	Average	93.6	77.0	84.0	0.98
	Standard deviation	1.5	1.9	1.5	0.07

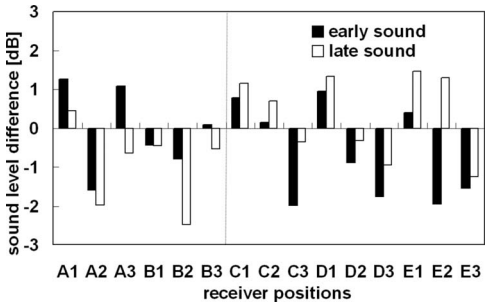


FIG. 7. Areal differences in energy level between scattered and specular sound fields: early reflection (3–80 ms) and the late part of reflection (80–∞ ms).

ber. The music source for the test was a 4 s violin motif. The power spectrum of the source signal indicated that the main components were above 500–4 kHz.

The convolved music stimuli were presented in a semianechoic chamber from two loudspeakers using the stereo dipole technique.[27,28] Figure 9 shows the experimental setup of loudspeakers. The speakers were located at a height of 1.2 m from floor. The distance between the speakers and subjects was 2 m and the angle between the speakers and the subjects was 19°. A cross-talk canceling filter was employed for precise binaural reproduction.

Two stereo impulse responses between the two channels of the dummy head and 2 loudspeakers were measured with segments of pseudorandom noise (MLS) in order to compute the respective inverse filters for cross-talk canceling according to the Nelson–Kirkeby method of frequency domain regularization.[28,29] Virtual impulse responses from the stereo dipole system were measured using a swept-sine signal convolved with the model hall impulse responses transformed into real scale at each seat. Four acoustical parameters of the virtual impulse response were compared with the model hall results, RT, EDT, C_{80}, and IACC. Results showed that differences in average values for the 1 and 2 kHz octave bands between two impulse responses are within the difference limens proposed by Bork[30] and Prodi and Velecka,[28] and just noticeable difference[25] (RT, 10% of the original value; EDT, 5% of the original value; C_{80}=1 dB; IACC=0.075). Cross-talk level was also measured with a white noise stereo signal (one channel at a time) to which the inverse filters were not

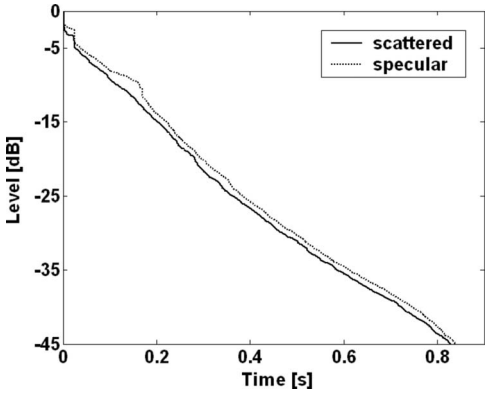


FIG. 8. Backward integration curves of scattered and specular sound at position A2.

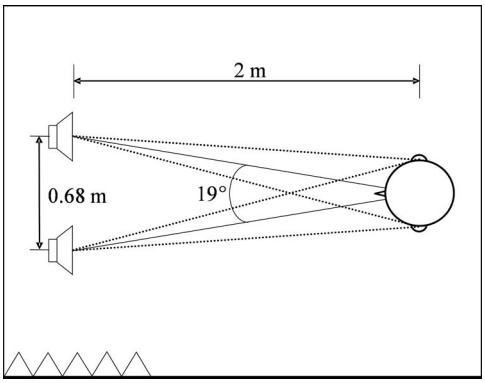


FIG. 9. Experimental setup of loudspeakers for subjective preference testing.

applied. The values of cross-talk were obtained from 1/3-octave-band level differences between left and right recorded channels. According to Ref. 28, 10 dB of the cross-talk level could be a suitable reference for indicating sufficient performance. Cross-talk levels for all pass white noise were 7.1 and 7.7 dB for the left and right channels, respectively. The left and right channel levels were 6.2 and 4.5 dB when those were calculated for frequency bands lower than 1 kHz. However, the cross-talk levels at frequency bands of 0.5–4 kHz were shown as 9.4 and 10.1 dB, which can be sufficient to eliminate the cross-talk. Also, a level of frequency range less than 500 Hz was more than 30 dB smaller than the level of 0.5–4 kHz where most of the sound energy of the music source is included.

2. Results

Figure 10 shows preferences of the scattered sound field at all 15 positions. All seats adjacent to the lateral wall (A3, B3, C3, D3, and E3) showed a high percentage of preference. A1, E1, and B2 also showed a high percentage of preference. Thus, the scattered sounds at these seats were highly preferred. However, at the seats at the center of the middle stalls (A2, C2, D2, and E2), the specular sound field was preferred.

Table III shows the correlation coefficients between the overall impression (preference) and subjective attributes of the scattered sound fields. As shown in Table III, the overall impression of scattered sound fields was correlated with loudness and reverberance. This means that subjects in this experiment prefered louder and more reverberant sound. There was a direct correlation between reverberance and en-

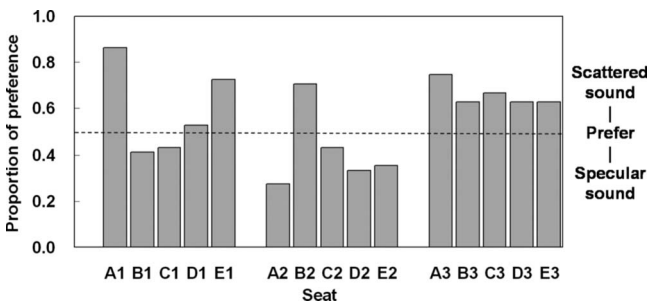


FIG. 10. Preference of scattered sound fields at all 15 positions in subjective experiment 1 in which 31 subjects participated.

TABLE III. Correlation coefficients between overall impression and subjective attributes for the scattered sound field.

	Overall impression	Loudness	Clarity	Reverberance	Envelopment	Apparent source width
Overall impression	···	0.65[b]	0.04	0.56[a]	0.31	0.12
Loudness		···	0.19	0.21	−0.21	−0.29
Clarity			···	−0.25	−0.43	−0.42
Reverberance				···	0.72[b]	0.62[b]
Envelopment					···	0.84[b]
Apparent source width						···

[a]$p<0.05$.
[b]$p<0.01$.

velopment. The results also showed that envelopment and apparent source width correlate highly with each other.

The relationship between the difference of the objective parameters and the overall impression was sought to determine the factors affecting acoustical attributes for scattered sound. The differences between the objective parameters were related to the scale value of the scattered sound (field) minus that of the specular sound (field). As a result, the overall impression of the scattered sound field was not significantly correlated with the objective parameter difference (with and without diffusers on the sidewalls).

B. Experiment 2

1. Method

In the previous auditory experiment, no significant relationship between the subjective preference of scattered sound fields and the objective parameters was found even though diffusers on the sidewalls affected acoustical parameters in the sound field. This is because subjects could not adequately perceive the difference in objective parameters with and without diffusers installed on the sidewalls.

In this experiment, 6 of the 15 seats used in the previous experiment were selected for an additional auditory test that focused on varying 1-IACC$_{E3}$. These seats consisted of those for which the 1-IACC$_{E3}$ increased (A1, A3, and B2) and decreased (D3, E2, and E3) the most with regard to distance from the source and sidewalls. Twelve impulse responses were used for the auditory test. The test was conducted using a paired-comparison method. Sixty-six sound pairs were presented to the subjects. Other details of this test were the same as in the previous experiment. The same 20 senior music subjects in the previous test also participated in this experiment. Each subject was asked to select the preferred sound of each sound pair. The scale values indicating the degree of subjective preference were calculated by applying the method of paired comparative judgment (Thurstone's case V)[31] and obtaining individual preference.[16]

2. Results

The subjective preference results were evaluated by a consistence test and an agreement test.[32] Both tests are useful tools for investigating subjective consistency and tendency from intrasubject and intersubject differences. Of the 20 subjects who participated in the test, 17 passed the consistency test. Also, these 17 subjects showed significant agreement ($p<0.01$), in spite of a low coefficient of agreement.

Figure 11 shows the scale values of sound fields with and without installed diffusers. In Fig. 11, scale values were calculated by the method of paired comparative judgment (Thurstone's case V) with data from the 17 subjects. As shown in Fig. 11, the subjects preferred the sound fields of A1, A3, and B2 to those of D3, E2, and E3. Seats A3, D3, and E3 which are adjacent to the sidewall were preferred in both experiments 1 and 2 when diffusers were installed. For seat A1, the difference in preference between scattered and specular sound fields was the largest among all seats in both experiments.

The correlation between the objective parameters and preference, and correlation among objective parameters are listed in Table IV. As shown in Table IV, preferences correlated ($p<0.05$) with RT and EDT significantly. The results also show that RT and EDT have a negative relationship with 1-IACC$_{E3}$.

Multiple regression analysis was used to determine the parameters which most greatly influenced subjective preference. All seven acoustical parameters (SPL, RT, EDT, C_{80}, ITDG, 1-IACC$_{E3}$, and 1-IACC$_{L3}$) were used as independent variables in the analysis. Stepwise, multiple regressions were conducted to obtain an optimal equation between all subjects and individual preferences, and objective parameters. The results show that the optimal equations from the analysis consist of SPL, EDT, or 1-IACC$_{L3}$. The significant standard-

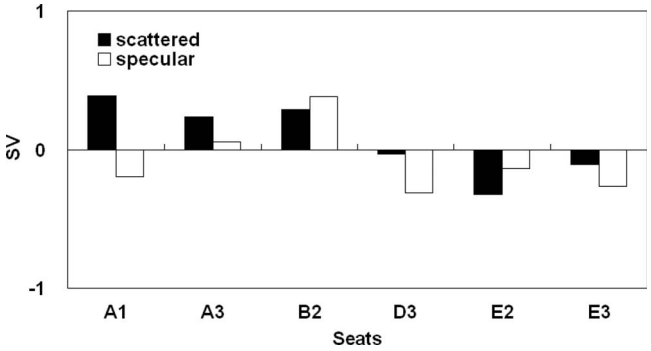


FIG. 11. Scale values of sound fields with and without installed diffusers in subjective experiment 2 in which 20 subjects participated.

TABLE IV. Correlations among the objective parameters and correlation between objective parameters and preference in experiment 2.

	Preference	SPL	RT	EDT	C_{80}	ITDG	1-IACC$_{E3}$	1-IACC$_{L3}$
Preference	⋯	0.46	0.69[a]	0.70[a]	−0.31	0.53	−0.49	0.51
SPL		⋯	0.59[a]	0.64	0.24	0.54	−0.57	0.08
RT			⋯	0.98[b]	−0.35	0.88[b]	−0.84[b]	0.26
EDT				⋯	−0.33	0.90[b]	−0.90[b]	0.28
C_{80}					⋯	−0.12	0.19	−0.29
ITDG						⋯	−0.91[b]	−0.04
1-IACC$_{E3}$							⋯	−0.15
1-IACC$_{L3}$								⋯

[a]$p<0.05$.
[b]$p<0.01$.

ized partial regression coefficients in Eq. (1) for all subjects and individuals are listed in Table V. No significant regression coefficient of parameters was obtained for individual preference of the five subjects.

$$\mathrm{SV}_{\mathrm{preference}} \approx a_1\mathrm{SPL} + a_2\mathrm{EDT} + a_3(1 - \mathrm{IACC}_{L3}) + c. \quad (1)$$

As shown in Table V, the dominant parameter in the determination of subjective preference was found to be EDT for all subjects. The parameters with the best fit in the optimal equation for all subjects were SPL and EDT, except for subjects E and H. Coefficients a_1 and a_2 have positive values in most cases, and thus larger sound levels and longer early decay times. These determine general preference of the sound fields. However, two subjects (I and J) did not prefer a louder and more reverberant sound field.

IV. DISCUSSION

The diffusers installed on the sidewalls reduced the amplitude of the first reflection because the sound energy of the incidence sound was dispersed in various directions, and the sound energy reaching the receiver points was decreased by the diffusers. These effects were mainly observed at the side positions where ITDG is shorter than 20 ms. However, at the rear receiver positions, the amplitude of the first reflection decreased and reflection with maximum amplitude was observed at the late parts of impulse responses.

TABLE V. Standardized regression coefficient of each acoustical parameter for all subjects and individuals.

	Regression coefficient			
Subjects	a_1	a_2	a_3	r
All	⋯	0.70[a]	⋯	0.70[a]
A	0.80[b]	⋯	⋯	0.80[b]
B	⋯	0.70[b]	⋯	0.69[b]
C	⋯	0.81[b]	⋯	0.81[b]
D	0.46[a]	0.57[b]	⋯	0.93[b]
E	⋯	0.59[b]	0.63[b]	0.90[a]
F	⋯	0.66[a]	⋯	0.65[b]
G	0.51[a]	0.52[a]	⋯	0.93[a]
H	⋯	⋯	0.66[a]	0.66[b]
I	−0.76[b]	⋯	⋯	0.76[b]
J	−0.55[a]	−0.45[a]	⋯	0.91[b]
K	0.79[b]	⋯	⋯	0.79[b]
L	⋯	0.74[b]	⋯	0.74[b]

[a]$p<0.05$.
[b]$p<0.01$.

The reduction of RT, EDT, and SPL was due to increased scattering (obvious from computer simulation models) and greater absorption resulting from the installation of the diffusers. C_{80} increased at the front areas (rows A and B) and decreased at the rear areas (rows D and E) showing that the difference of early sound energy from the installed diffusers was higher than late-reflection sound energy at the front seats (A1-B3). However, opposite results were observed at the rear seats (C1-E3). The diffusing elements (hemisphere- and polygon-shaped diffusers) on the walls, soffit, and balcony fronts in the scale model led to more sound being reflected toward the front and less to the rear. Barron and Lee[33] also observed the increase of the early energy levels at the front seats from highly diffusing ceilings in concert halls. Consequently, from the areal values of acoustical parameters and the early reflection pattern, it was found that effects of diffusers on the sound fields vary according to receiver positions and diffuser locations.

When considering the distribution of SPL and RT throughout the opera house, the measurement results of acoustical parameters indicate that there is no significant difference in total sound energy between with and without diffusers installed on the sidewalls. However, the late sound energy in the diffuse sound field was more evenly distributed due to the presence of the diffusers. Conversely, the average RT value was slightly decreased by diffusers, but there was no significant difference in standard deviation of RT between with and without diffusers. The present results of SPL and RT variations in the specular and scattered sound fields are somewhat different from the results of previous studies,[15,34] in which the variation of reverberation time distribution was reduced by the addition of diffusing panels but the sound levels were not affected.

In the subjective experiment 1, it was found that overall impression (preference) of scattered sound fields is highly correlated with loudness and reverberance, but was not significantly correlated with the objective parameter difference between sidewalls with or without diffusers. In subjective experiment 2, EDT and SPL of acoustical parameters have the greatest effects on subjective preference for scattered and

specular sounds. Considering the results of subjective experiments 1 and 2, EDT and SPL were determined to be dominant acoustical parameters describing subjective preference for scattered sound field. However, in the objective measurements, EDT and SPL decreased at most of the front seats when diffusers were installed on the wall. This observation illustrates that if too many diffusers are installed on the wall, the audience may not prefer the scattered sound field because the use of diffusers causes both EDT and SPL to decrease.

The results of the present investigation were different from those of a previous study[16] in which the IACC was found to be a significant parameter for describing "subjective diffuseness." In this study, the subjects preferred sound sources with a lower 1-IACC$_{E3}$ value, whereas subjective preference tests (in which only IACC varies) for sound fields in a previous study showed that the scale value of preference increased as IACC decreased. This difference seems to be due to the fact that the correlation between EDT and 1-IACC$_{E3}$ was high ($r=-0.89$, $p<0.05$) in the present scale model experiment. Another reason for the difference between the two studies is that differences in IACC values were not easily perceived in the present experiment. Therefore, the subjects in this study determined their subjective preference of sound fields on the basis of reverberance and loudness. Hotehama *et al.*[35] also used spatial factors to find out the subjective accuracy for source location when there is enough perceptual difference in the factors. Their findings are supported by both temporal and spatial factors. However, the target of their investigation is different from the present study, which is more concerned with subtle differences caused by diffuse sound fields.

The limitations in spatial hearing by scale model head without pinna[22,24] in the present study might affect subjective preferences. However, the main purpose of this study is to make relative comparisons between conditions with and without diffusers while other conditions remained controlled. In addition, the comparison test with actual dummy showed that scale model head has reasonable quality for measure of spatial impression. The scale model hall measurement using scale model head was found to well represent the difference in IACC between receiver positions in actual hall measurement. The reproduction system in the subjective tests showed validity of the simulated sounds. There was another limitation in the present study; the sampling rate of the music source was adjusted from 192 to 19.2 kHz for subjective tests. However, the sampling rate is high enough to evaluate the 1 and 2 kHz octave band, which was assumed as the effective frequency range in terms of scattering performance.

Therefore, the data from the present binaural measurement and reproduction system are regarded as reasonable on the basis of physical analysis. However, psychoacoustical investigation of scale dummy head such as rate of in-head-localization and localization errors, difference of spaciousness, and room impression should be further conducted for validity of dummy head. In this study, late lateral energy[36] (LG) and spatially balanced center time[37] (SBT) relevant to listener environment (LEV) were not measured since there was no suitable device for scale model measurement. Because LEV may be relevant to subjective effects of diffusion in a hall, there is a need for investigation of the variations of LG and SBTs with or without installed diffusers.

V. CONCLUSIONS

Scattered sound fields in an opera house were objectively and subjectively evaluated through acoustical measurements in a 1:10 scale model and through subjective preference tests. The objective evaluations consisted of determining the effects of diffusers on the walls close to the proscenium, sidewalls, and soffit of the side balcony. The subjective evaluations employed auditory experiments to determine subjective preference for scattered or specular sound fields with and without diffusers, and to determine the effective parameters for describing preference of the sound fields. All parameter values except for IACC were given as values derived by averaging the results at 1 and 2 kHz octave bands according to the effective frequencies of the installed diffusers.

When diffusers were installed, the SPL of the first reflection was reduced by an average of 5 dB in amplitude at rear receiver positions and near the sidewall. This was because the sound energy of incidence sound was dispersed and the sound energy reaching the receiver positions was decreased by the diffusers. RT and EDT decreased as well due to the increased scattering and greater absorption at most seats. C_{80} and 1-IACC$_{E3}$ increased at the front seats, while the values deceased at the rear seats near installed diffusers. These results showed that the effects of scattered sound on acoustical parameters in the sound field depend on the receiver position and diffuser location.

The subjective preference test revealed that preference of scattered sound fields correlates highly with loudness and reverberance. It was also found that EDT and SPL are the dominant acoustical parameters that describe preference for scattered sound field under the experimental conditions employed in this study. The results of the subjective and objective evaluations indicated that that if too many diffusers are installed on the wall, the audience may not prefer the scattered sound field because several diffusers cause EDT and SPL to be reduced.

Temporal and spatial parameters including late lateral energy, which were not measured in this study, should be further investigated to represent the degree of scattering in a sound field. Since the effects of diffuse surfaces may vary depending on architectural variables such as room shape, dimension, and location of the diffuser, future studies should focus on *in situ* measurement of the scattering coefficient.

ACKNOWLEDGMENTS

We are grateful for the assistance we have received from several researchers; however, in particular, we would like to thank Dr. Shin-ichi Sato for help on the data analysis and the reproduction system for the subjective test. We also gratefully acknowledge the review and comments of Dr. Mike Barron, Professor Michael Vorländer and Professor Yoichi Ando.

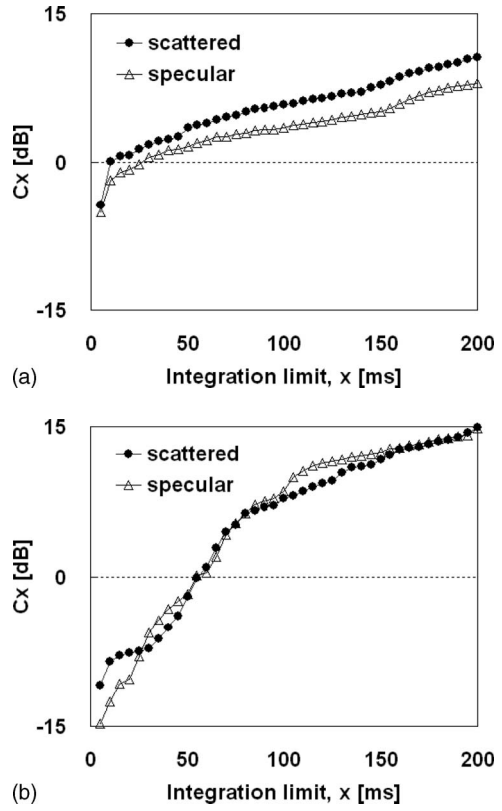


FIG. 12. Typical examples of C_x. (a) Seat B2 and (b) seat E1.

APPENDIX A: DISCUSSION ABOUT INTEGRATION LIMIT THAT DIVIDES THE EARLY AND LATE PARTS OF IMPULSE RESPONSES

There is a discussion for the integration limit that divides the early and late parts of the impulse response. We investigated the early to late energy ratio C_x ($10 \leq x \leq 200$ ms) proposed by Marshall.[38] Two typical examples are shown in Fig. 12. At the frontal seats (A1-D1), C_x for the diffuse sound field is larger than that for the specular sound field regardless of the integration limit. On the contrary, at the rear seats (D2-E3), the relationship between C_x for the specular and diffuse sound fields change according to x. Table VI shows the correlation among C_x ($10 \leq x \leq 100$ ms). In the cases of x less than 60 ms, C_x is not correlated with C_x more than 70 or 80 ms significantly ($p > 0.01$). In contrast, C_x of 60, 70, and 80 ms is highly correlated with C_x over 70 or 80 ms significantly ($p < 0.01$). The correlations among C_x ($100 \leq x \leq 200$ ms) are always more than 0.90 ($p < 0.01$) for all cases. These results are more clearly seen in the data of C_x more than 100 ms. C_x of 60, 70, and 80 ms has a significantly high correlation with C_x of 100 ms ($p < 0.01$). Consequently, 60–80 ms can be considered as the time boundary to divide early and late parts of the impulse response.

Similarly, 80 ms has been used for the integration limit that divides the early and the later parts of the impulse response, especially for clarity measurement in the ISO standard.[25] Barron[39] found that the spatial impression can be measured by the ratio of lateral to nonlateral energy arriving within the first 80 ms. Reichardt *et al.*[40] introduced clarity, C_{80}, for the music signal as well as C_{50} for speech signal. Barron and Marshall[41] found that 80 ms corresponds roughly to the area of transition between the effects of early reflections and those of reverberation. It has the unfavorable effect that in critical cases a small change in the arrival time of a strong reflection may result in a significant change in C_{80}. Therefore, the center time, T_s, which does not involve any delay limit, was defined as the first moment of the squared impulse responses. The correlation between C_x and T_s was also investigated (see Table VI); T_s showed a high correlation with C_{60}-C_{80}.

APPENDIX B: INTERAURAL CROSS-CORRELATION COEFFICIENT (IACC) AND ITS PARAMETERS

There is also discussion about the frequency ranges of IACC to be related to spatial impression and the integration limit to divide the impulse response into the early and the later parts. Barron and Marshall[41] found that the broadening effect is greatly diminished for frequencies in the 4 kHz range. Blauert[24] confirmed less contribution of the 4 kHz band to the spatial impression of music signal. Okano *et al.*[42] found that 500–4 kHz octave bands are equally important in determining the apparent source width (AWS) Hidaka *et al.*[20]

TABLE VI. Correlation between C_x ($10 \leq x \leq 100$ ms, average for 1 and 2 kHz) for 30 cases (15 receiver positions for the specular and diffuse sound fields). Correlation with T_s is also indicated.

x (ms) and T_s										
10	20	30	40	50	60	70	80	90	100	T_s
10	0.91[a]	0.87[a]	0.82[a]	0.82[a]	0.68[a]	0.39	0.06	−0.06	−0.13	−0.46
20	⋯	0.93[a]	0.90[a]	0.87[a]	0.74[a]	0.45	0.11	0.00	−0.07	−0.49[a]
30		⋯	0.97[a]	0.91[a]	0.80[a]	0.54[a]	0.24	0.13	0.04	−0.49[a]
40			⋯	0.94[a]	0.85[a]	0.60[a]	0.30	0.19	0.10	−0.51[a]
50				⋯	0.94[a]	0.72[a]	0.44	0.34	0.25	−0.59[a]
60					⋯	0.88[a]	0.66[a]	0.57[a]	0.50[a]	−0.70[a]
70						⋯	0.90[a]	0.84[a]	0.79[a]	−0.76[a]
80							⋯	0.98[a]	0.95[a]	−0.64[a]
90								⋯	0.98[a]	−0.60[a]
100									⋯	−0.59[a]
T_s										⋯

[a] $p < 0.01$.

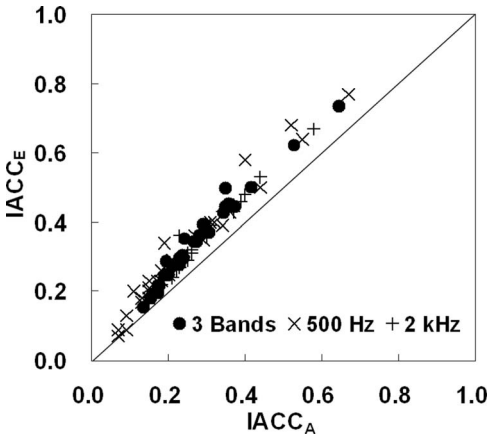


FIG. 13. Relationship between the values of the $IACC_E$ and IACC. Three bands are 0.5, 1, and 2 kHz octave band.

proposed an average of a simple summation of the three bands with center frequencies of 0.5, 1, and 2 kHz to determine the most sensitive form of IACC. They considered the fact that the level of the music at the 4 kHz octave band for a typical symphonic piece is 15 dB lower than the levels in the 0.5–2 kHz bands.[43] They found that the variation of IACC at 125 and 250 Hz, according to the different positions in a hall is less sensitive than at the higher frequency bands. They also defined $IACC_{E3}$ and $IACC_{L3}$, which are related to the subjective ranking of concert hall quality and to the degree of sound diffusion, respectively.

Hidaka *et al.*[20] also investigated the integration limits of IACC to determine the $IACC_E$'s for t equal to 50, 80, 100, and 200 ms. Their results showed that $IACC_E$'s for 0–80 and 0–200 ms have sufficiently high correlations to permit selection of any time in this range to divide early from late. Later Beranek[1] used the term "binaural quality index (BQI)" measured by 1-$IACC_{E3}$ as an index that has the best correlation with the subjective rank orders for concert halls and opera houses. As shown in Fig. 13, the integration limit 80 ms has wider range in IACC of all measured receiver positions in the present study; thus it is more objective and useful in distinguishing the sound fields.

[1]L. L. Beranek, *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York, 2004).
[2]M. Barron, "The subjective effects of first reflections in concert halls: The need for lateral reflections," J. Sound Vib. **15**, 475–494 (1971).
[3]P. D'Antonio and T. J. Cox, "Diffusor application in rooms," Appl. Acoust. **60**, 113–142 (2000).
[4]T. J. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers: Theory, Design and Application* (Spon, New York, 2004).
[5]P. D'Antonio and T. J. Cox, "Two decades of sound diffuser design and development. Part 1: Applications and design," J. Audio Eng. Soc. **46**, 955–976 (1998).
[6]T. J. Cox, B.-I. L. Dalenback, P. D'Antonio, J. J. Embrechts, J. Y. Jeon, E. Mommertz, and M. Vorländer, "A tutorial on scattering and diffusion coefficients for room acoustic surfaces," Acust. Acta Acust. **92**, 1–15 (2006).
[7]M. Hodgson, "Evidence of diffuse surface reflections in rooms," J. Acoust. Soc. Am. **89**, 765–771 (1991).
[8]B.-I. L. Dalenbäck, M. Kleiner, and P. Svensson, "A macroscopic view of diffuse reflection," J. Audio Eng. Soc. **42**, 793–807 (1994).
[9]Y. W. Lam, "A comparison of three diffuse reflection modeling methods used in room acoustics computer models," J. Acoust. Soc. Am. **100**, 2181–2192 (1996).
[10]M. Vorländer and E. Mommertz, "Definition and measurement of random-incidence scattering coefficients," Appl. Acoust. **60**, 187–199 (2000).
[11]J. Y. Jeon, S. C. Lee, and M. Vorländer, "Development of scattering surfaces for concert halls," Appl. Acoust. **65**, 341–355 (2004).
[12]ISO 17497-1, "Acoustics—Sound scattering properties of surfaces. Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room," 2004.
[13]AES-4id-2001, "AES information document for room acoustics and sound reinforcement systems characterization and measurement of surface scattering uniformity," J. Audio Eng. Soc. **49**, 149–165 (2001).
[14]Y. Suzumura, M. Sakurai, Y. Ando, I. Yamamoto, T. Iizuka, and M. Oowaki, "An evaluation of the effects of scattered reflections in a sound field," J. Sound Vib. **232**, 303–308 (2000).
[15]S. Chiles and M. Barron, "Sound level distribution and scatter in proportionate spaces," J. Acoust. Soc. Am. **116**, 1585–1595 (2004).
[16]Y. Ando, *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners* (AIP, New York and Springer-Verlag, New York, 1998).
[17]R. R. Torres, M. Kleiner, and B.-I. Dalenback, "Audibility of 'diffusion' in room acoustics auralization: An initial investigation," Acust. Acta Acust. **86**, 919–927 (2000).
[18]D. Takahashi and R. Takahashi, "Sound fields and subjective effects of scattering by periodic-type diffusers," J. Sound Vib. **258**, 487–497 (2002).
[19]C. H. Haan and F. R. Fricke, "An evaluation of the importance of surface diffusivity in concert halls," Appl. Acoust. **51**, 53–69 (1997).
[20]T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," J. Acoust. Soc. Am. **98**, 988–1007 (1995).
[21]ISO 354, "Acoustics—Measurement of sound absorption in a reverberation room," 2003.
[22]N. Xiang and J. Blauert, "A Miniature Dummy Head for Binaural Evaluation of Tenth-Scale Acoustic Models," Appl. Acoust. **33**, 123–140 (1991).
[23]N. Xiang and J. Blauert, "Binaural Scale Modeling for Auralization and Prediction of Acoustics in Auditoria," Appl. Acoust. **38**, 267–290 (1993).
[24]J. Blauert, *Spatial Hearing*, (MIT, Cambridge, MA, 1983).
[25]ISO/CD 3382-1, "Acoustics—Measurement of the reverberation time—Part 1: Performance spaces," 2005.
[26]M. Barron, "Subjective study of British symphony concert halls," Acustica **66**, 1–14 (1988).
[27]O. Kirkeby, P. A. Nelson, and H. Hamada, "The 'stereo dipole'—a virtual source imaging system using two closely spaced loudspeakers," J. Audio Eng. Soc. **46**, 387–395 (1998).
[28]N. Prodi and S. Velecka, "The evaluation of binaural playback systems for virtual sound fields," Appl. Acoust. **64**, 147–161 (2003).
[29]O. Kirkeby and P. A. Nelson, "Digital filter design for virtual source imaging systems," J. Audio Eng. Soc. **46**, 387–395 (1998).
[30]I. Bork, "Comparison of room simulation software—the 2nd round robin on room acoustical computer simulation," Acustica **86**, 943–956 (2000).
[31]L. L. Thurstone, "A law of comparative judgment," Psychol. Rev. **34**, 273–286, (1927).
[32]F. Mosteller, "Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed," Psychometrika **16**, 207–218 (1951).
[33]M. Barron and L.-J. Lee, "Energy relations in concert auditoriums," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[34]T. E. Bartel and E. B. Magrab, "Studies on the spatial variation of decaying sound fields," J. Acoust. Soc. Am. **63**, 1841–1850 (1978).
[35]T. Hotehama, S. Sato, and Y. Ando, "Dissimilarity judgments in relation to temporal and spatial factors for the sound fields in an existing hall," J. Sound Vib. **258**, 429–441 (2002).
[36]J. S. Bradley and G. A. Soulodre, "Objective measures of listener envelopment," J. Acoust. Soc. Am. **98**, 2590–2597 (1995).
[37]T. Hanyu and S. Kimura, "A new objective measure for evaluation of listener envelopment focusing on the spatial balance of reflections," Appl. Acoust. **62**, 155–184 (2001).
[38]L. G. Marshall, "An analysis procedure for room acoustics and sound amplification system based on the early-to-late sound energy ratio," J. Audio Eng. Soc. **44**, 373–381 (1996).
[39]M. Barron, "The subjective effects of first reflections in concert halls-the need for lateral reflections," J. Sound Vib. **15**, 475–494 (1971).
[40]W. Reichardt, O. Abdel Alim, and W. Schmidt, "Definition und Messgmndlage eines objectiven Masses zur Emittlung der Grenze zwischen brauchbarer unbrauchbarer Durchssichtigkeit bei Musikdarbietungen (Definition and basis of making an objective evaluation between useful and useless clarity defining musical performances)," Acustica **32**, 126–137

(1975).
[41]M. Barron and A. H. Marshall, "Spatial impression due to early reflections in concert halls: The of a physical measure lateral derivation," J. Sound Vib. **77**, 211–232 (1981).
[42]T. Okano, T. Hidaka, and L. L. Beranek, "Relations between the apparent source width (ASW) of the sound field in a concert hall and its sound pressure level at low frequencies (GL), and its inter-aural cross correlation coefficient (IACC)," J. Acoust. Soc. Am. **96**, 3268 (1994).
[43]T. Okano, L. L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls," J. Acoust. Soc. Am. **104**, 255–265 (1998).

Sound absorption by clamped poroelastic plates

H. Aygun[a] and K. Attenborough[b]
Department of Engineering, University of Hull, Cottingham Road, Hull HU6 7RX, United Kingdom

(Received 27 July 2007; revised 14 May 2008; accepted 2 June 2008)

Measurements and predictions have been made of the absorption coefficient and the surface acoustic impedance of poroelastic plates clamped in a large impedance tube and separated from the rigid termination by an air gap. The measured and predicted absorption coefficient and surface impedance spectra exhibit low frequency peaks. The peak frequencies observed in the absorption coefficient are close to those predicted and measured in the deflection spectra of the clamped poroelastic plates. The influences of the rigidity of the clamping conditions and the width of the air gap have been investigated. Both influences are found to be important. Increasing the rigidity of clamping reduces the low frequency absorption peaks compared with those measured for simply supported plates or plates in an intermediate clamping condition. Results for a closed cell foam plate and for two open cell foam plates made from recycled materials are presented. For identical clamping conditions and width of air gap, the results for the different materials differ as a consequence mainly of their different elasticity, thickness, and cell structure.
 [DOI: 10.1121/1.2951586]

PACS number(s): 43.55.Ev, 43.20.Tb, 43.20.Gp, 43.40.Dx [NX] Pages: 1550–1556

I. INTRODUCTION

Sound absorbers including porous materials are used widely for noise control. The most widely exploited and acknowledged absorption mechanism in porous materials is viscous friction due to relative motion between solid and fluid. If the frame of the porous material is viscoelastic then other dissipative mechanisms are possible. To obtain good low frequency absorption, say, below 300 Hz, exploiting the viscous friction mechanism alone may require an unacceptably large thickness of material. In designing passive sound absorbers based on porous materials or membranes, it is known that the presence of a backing air gap with properly chosen dimensions can enhance their low frequency performance. If a porous plate is elastic then a backing air cavity will allow bending modes which could lead to significant absorption at relatively low frequencies. Indeed such low frequency absorption maxima have been observed for configurations consisting of clamped poroelastic plates with an air gap behind them[1]. It has been found that the coupling between airborne sound and bending vibrations in the plates is increased if the flow resistivity is high. Depending on its thickness, the presence of an air gap has been thought previously to favor or hinder structural vibration but not to change the resonance frequencies which are considered to be solely dependent on the plate's mechanical properties, the clamping conditions, and the microstructural parameters (porosity, tortuosity, and flow resistivity)[1].

Takahashi and Tanaka[2] developed an analytical model by introducing flow continuity at the plate surface in a spatially mean sense and air-solid interaction within the plate material. However, their formulation does not allow for clamping at the boundaries of the plates. Horoshenkov *et al.*[3,4] formulated a model for the acoustic response of elastic porous plates and the effect of plate vibration. They have used a coupling condition between the plate and surrounding air at the plate surface and have shown that the low frequency resonant peaks observed in the surface impedance of the clamped plates with air backing are related to modes of vibration of the plates. Leclaire *et al.*[5] presented a variational method for solving the plate equations for different porous elastic plates with the four edges clamped but have neglected the effect of fluid loading on the plate vibration and have not taken into account external forces other than the excitation terms. Recently Aygun *et al.*[6,7] included the effects of fluid loading on the vibration of clamped, rectangular porous elastic plates by solving the governing equations for flexural vibration. Attenborough and Aygun[8] demonstrated the existence of low frequency absorption coefficient resonances in configurations consisting of clamped poroelastic plates with an air gap between the plates and a rigid termination. As well as the size of the plate and the corresponding clamping conditions, the width of the rear air gap is found to alter the absorption peak magnitudes and frequencies. This paper reports these results in greater detail together with additional data and predictions.

Measurements have been made of the absorption coefficient and the acoustic surface impedance of a closed cell [polyvinylidene fluoride (PVDF)] foam plate and two open cell foam plates made from recycled materials clamped in a large impedance tube with concrete walls. The tube has been designed to enable measurements between 50 and 300 Hz. The measured absorption coefficients have been compared with those predicted for fluid loaded, clamped, rectangular porous elastic plates. To investigate the importance of porosity, comparative predictions have been made for nonporous plates with the same elasticity parameters as the poroelastic plates. To emphasize the importance of the elasticity of the clamped porous plate, calculations have been made also of

[a]Electronic mail: h.aygun@hull.ac.uk
[b]Electronic mail: k.attenborough@hull.ac.uk

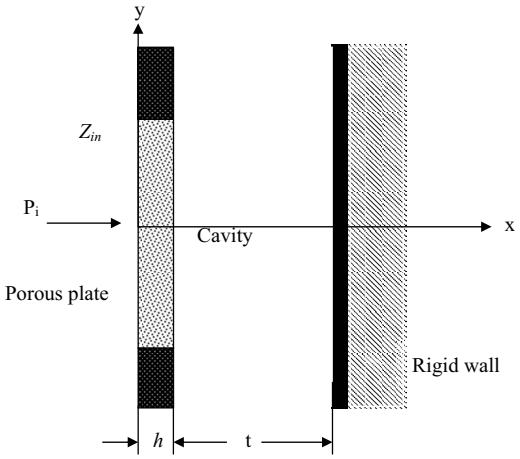


FIG. 1. Schematic of a poroelastic plate in a rigid baffle separated from a rigid wall by an air gap.

the surface impedance and normal incidence absorption coefficient of a rigid-porous plate with the same pore-related properties as those of the porous elastic plate. To investigate further the extent to which the surface impedance resonances are related to plate vibration, measurements have been made of the deflection spectra while the plates were clamped in the impedance tube.

II. THEORY

Consider a thin, fluid loaded, poroelastic plate of infinite length in the z-direction and a finite width, L, in the y-direction, and thickness h, as shown in Fig. 1.

The poroelastic plate is assumed to be clamped at its edges in the y-direction and separated from the rigid wall by an infinitely long air cavity of thickness t which allows bending vibration. The plate is subjected to a normal incidence plane wave, $P_i = e^{ik_0 x}$, where k_0 is the wave number in air and harmonic time dependence is understood. The air-filled cavity has an impedance given by $Z_a = \rho_0 c_0 \coth(-ik_0 t)$, where ρ_0 is the air density and c_0 is the speed of sound in air. The normal incidence surface acoustic impedance at the front of the poroelastic plate is given by[3]

$$Z_{in} = Z_p \frac{Z_a + Z_p \tanh(-ikh)}{Z_p + Z_a \tanh(-ikh)}, \tag{1}$$

where Z_p and k are the characteristic impedance and the wave number, respectively. Strictly these should be calculated for the slow wave in the poroelastic plate but here they have been calculated for a rigid-porous medium, with the relevant parameter values in Table I, by using the Johnson–Allard model.[9]

The averaged normalized surface impedance of the plate is given by[3]

$$\langle Z \rangle(\omega) = \frac{\int_0^L p_s(z)dz}{\rho_0 c_0 \int_0^L (-i\omega w_p(z) + p_s(z)/Z_{in})dz}, \tag{2}$$

where L is the width of the plate, ω is the angular frequency, p_s is the acoustic pressure on the front of the plate, and w_p is the displacement of the fluid loaded, finite, thin, clamped poroelastic plate which is given by Refs. 6 and 7 Although

TABLE I. Measured and deduced (*) properties of the poroelastic plates.

	YB10 foam plate	PVDF foam plate	Black foam plate
Lx (m) in x direction	0.5	0.5	0.6
Ly (m) in y direction	0.5	0.5	0.6
Thickness (m)	0.0107	0.025	0.021
Density ρ (kg/m^3)	353	25.15	223
Young's modulus (Pa)	2.1×10^7	254 600*	2.46×10^6
Loss factor	0.1	0.15*	0.35
Poisson ratio ν	0.35	0.35*	0.3
Porosity ϕ	0.69	0.2	0.75
Tortuosity τ_∞	1.2	1*	1.85
Permeability κ (m^2)	2.7×10^{-10}	3.24×10^{-11}	3.92×10^{-10}
Flow resistivity N s m^{-4}	68 111	567 901	46 933

Eq. (2) assumes that the plate is of infinite length, the calculations of w_p assume some form of clamping at all four edges and have been made using

$$w_p(z) = \sum_{m=1}^{\infty} W_m^p Z_m(z), \tag{3}$$

where $Z_m(z)$ are the beam functions and W_m^p is the plate lateral displacement. Expressions for W_m^p are given elsewhere.[6,7]

The normal incidence sound absorption coefficient is given by

$$\alpha(\omega) = 1 - \left| \frac{\langle Z \rangle - 1}{\langle Z \rangle + 1} \right|^2. \tag{4}$$

III. MEASUREMENTS

Measurements have been carried out in a large 6 m long rectangular impedance tube with an internal cross section of 50×50 cm^2 and external cross section of 65×65 cm^2 (see Fig. 2). The center-to-center spacing between two microphones is 20 cm. The distance from the test sample to the center of the nearest microphone is 50 cm. A standard procedure[10] has been followed during the measurements.

The acoustic sound field was created by a loudspeaker fed by a noise generator and a 600 W power amplifier. Two $\frac{1}{2}$ in. B&K microphones were mounted into microphone grid at positions along the length of the impedance tube. Each microphone grid was sealed tight to its housing. The microphones were supported by two preamplifiers which were

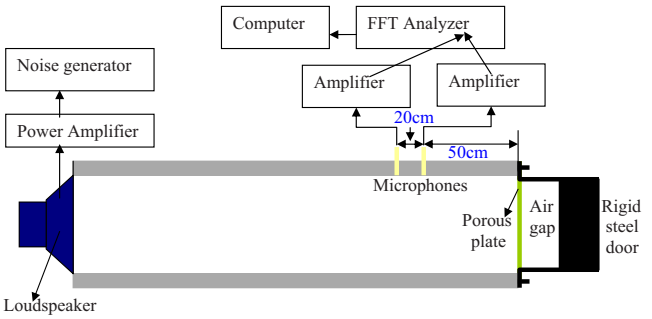


FIG. 2. (Color online) Impedance tube arrangement (Ref. 7).

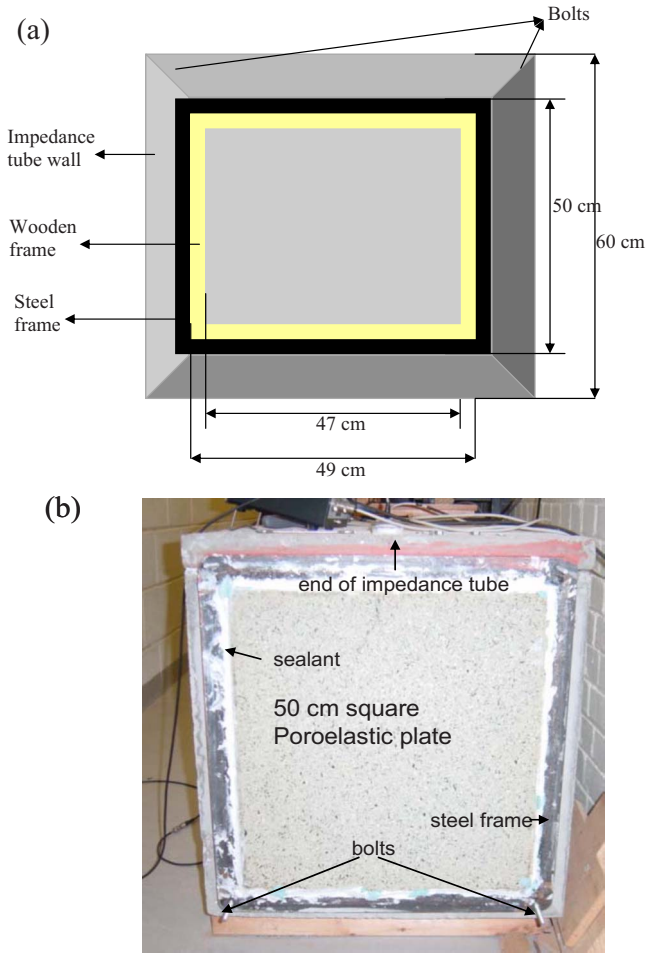


FIG. 3. (Color online) (a) A diagram of the end of the impedance tube and the plate clamping arrangement. (b) Photograph showing a 50 cm square YB10 plate mounted at the end of the impedance tube.

connected to two B&K measuring amplifiers. The outputs of the measuring amplifiers were fed to a two-channel fast Fourier transform analyzer which was connected to a computer for data analysis. The porous plate was separated from rigid steel door at the end of the tube by an air gap to allow bending vibration (see Fig. 2). The edges of the porous elastic plates being tested were clamped in a steel frame fitted inside the impedance tube. The edges of the porous elastic plate were compressed between the end of the impedance tube and the rigid backing system by using four bolts. The edges of 60 cm square plates were further secured in an additional wooden frame (see Fig. 3) which reduced the exposed area of the plates to 47×47 cm^2, i.e., the distance between the clamped edges of the porous elastic plate was 47 cm. Subsequently this condition is called "rigid" clamping (CCCC).

To achieve "simply supported" boundary conditions (denoted by SSSS in the figures), the smaller (50 cm square) plates were glued to the edges of the steel frame rather than clamped. Any gaps between the plate and frame were sealed by using double sided tape and sealant. A third type of clamping condition, subsequently described as "intermediate," was obtained with the 50 cm square plates by gluing them to the edges of the steel frame. Although the edges of the plates were compressed between the end of the impedance tube and the rigid backing using the bolts, the smaller plate size means that the clamping was not as tight as was the case for the "rigidly clamped" conditions.

As well as different sizes of two types of plate, three different types of plate materials have been used in the impedance tube measurements. The PVDF foam was 25 mm thick and predominantly closed cell. YB10 and Black plates, 10.7 and 21 mm thick, respectively, were open cell foams fabricated from recycled car dashboards. The measured and deduced properties of the poroelastic plates are given in Table I.

The listed properties of the recycled foam plates (Black and YB10) were measured at the University of Bradford[7]. The permeability of the PVDF foam has been calculated from the measured flow resistivity (567 700 N s m^{-4}). The elastic modulus of the PVDF Foam has been deduced from vibration data. Values of the loss factor, Poisson's ratio, and tortuosity values for the PVDF foam have been deduced by fitting the surface impedance data. The deduced properties are distinguished by asterisks in Table I.

IV. RESULTS OF MEASUREMENTS

The surface acoustic impedance $\langle Z\rangle$ at the front surface of the poroelastic plate is calculated from measurements by using

$$\langle Z\rangle = (1+R_c)/(1-R_c). \tag{5}$$

R_c, the complex reflection coefficient, is given by[10]

$$R_c = \frac{H-e^{iks}}{e^{-iks}-H}e^{j2k(l+s)}, \tag{6}$$

where l is the distance from the test sample to the center of the nearest microphone, s is the center-to-center spacing between microphones, and H is the measured transfer function of the two microphone signals corrected for microphone response mismatch.

The corresponding normal incidence absorption coefficient is obtained from

$$\alpha = 1-|R_c|^2. \tag{7}$$

A. Effects of changing clamping conditions and air gap width

The measured impedance and absorption coefficient spectra obtained with 50 and 60 cm square YB10 plates and a backing air gap width of 60 cm are shown in Figs. 4 and 5, respectively. The absorption coefficient spectra show useful low frequency peaks. The different sizes of plates result in different edge clamping conditions. As remarked earlier, it was not possible to obtain as rigidly clamped conditions with the smaller YB10 plate (50×50 cm^2) as was possible when using the larger plates. According to these data, the magnitudes and frequencies of the surface acoustic impedance and corresponding absorption coefficient peaks are significantly affected by the rigidity of the clamping conditions. The less rigid, intermediate clamping results in higher absorption peaks.

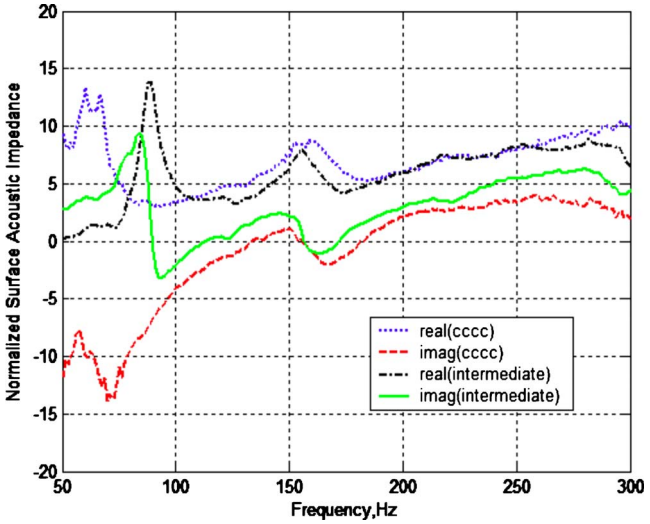


FIG. 4. (Color online) A comparison of the measured normalized surface acoustic impedance spectra for 60 and 50 cm square YB10 plates in front of a 60 mm wide air gap. The corresponding clamping conditions are rigidly clamped (CCCC) and "intermediate," respectively.

Figure 6 compares the measured surface acoustic impedances of 50 cm square YB10 plates, subject to intermediate clamping conditions but with 60 and 80 mm air gap widths, respectively. Figure 7 shows the corresponding absorption coefficient spectra. The larger air gap width results in a larger peak in the absorption coefficient near 100 Hz and this is shifted to lower frequency. On the other hand, the peak near 170 Hz is reduced in magnitude with the larger air gap width but there is little change in its frequency.

The effects of air gap width and clamping conditions have been investigated also with 60 cm square Black plates. Figure 8 shows data obtained with air gap widths of 60 and 80 mm under simply supported conditions and an air gap width of 60 mm under rigidly clamped conditions. With a 60 mm wide air gap, rigid clamping seems to eliminate the absorption resonance between 100 and 150 Hz compared with the simply supported condition. With the simply supported condition, increasing the width of the air gap causes

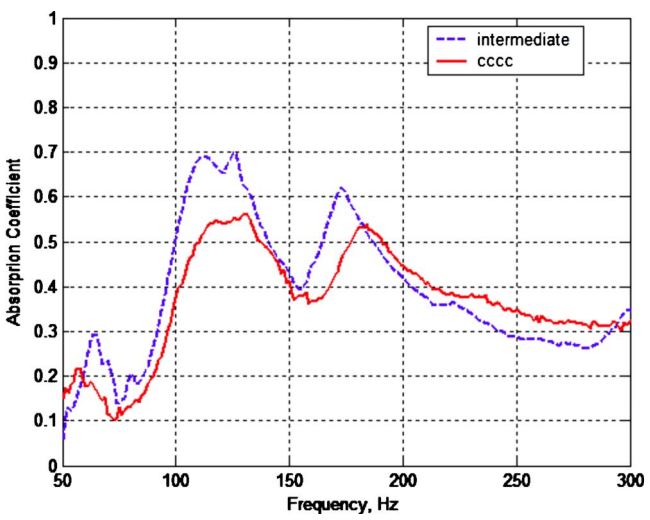


FIG. 5. (Color online) Measured absorption coefficient spectra corresponding to Fig. 4 (Ref. 7).

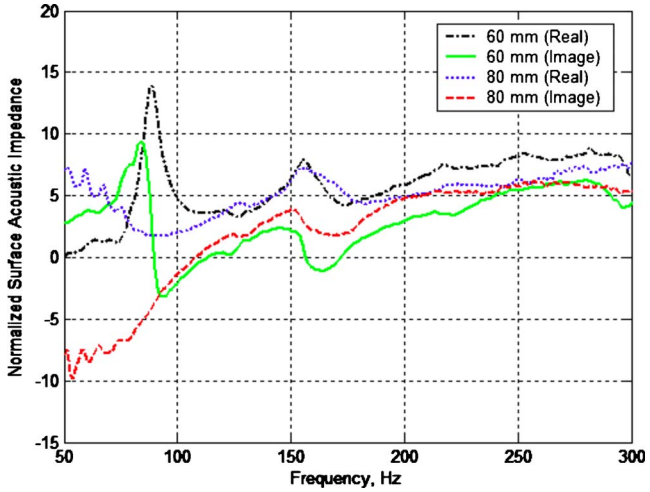


FIG. 6. (Color online) A comparison of measured normalized surface acoustic impedance spectra for the YB10 plate with 60 and 80 mm air gaps and intermediate clamping conditions.

this resonance to increase in magnitude. The differences between curves in Fig. 8 are because of the depth of the air gap and different boundary conditions.

B. Comparative performance of different plate materials

Even with simply supported boundary conditions, the low frequency absorption peaks obtained with the Black plate are significantly less pronounced than those obtained with YB10 plate. The YB10 plate is much stiffer than the Black plate. Moreover the thickness of the Black plate is almost double that of the YB10 plate.

The measured real and imaginary parts of the surface acoustic impedance of 50×50 cm^2 PVDF foam and YB10 foam plates, subject to intermediate clamping conditions, are shown in Fig. 9. The corresponding absorption coefficient spectra are compared in Fig. 10.

The measured surface impedance spectra for the YB10 foam plate exhibit a more pronounced resonant structure

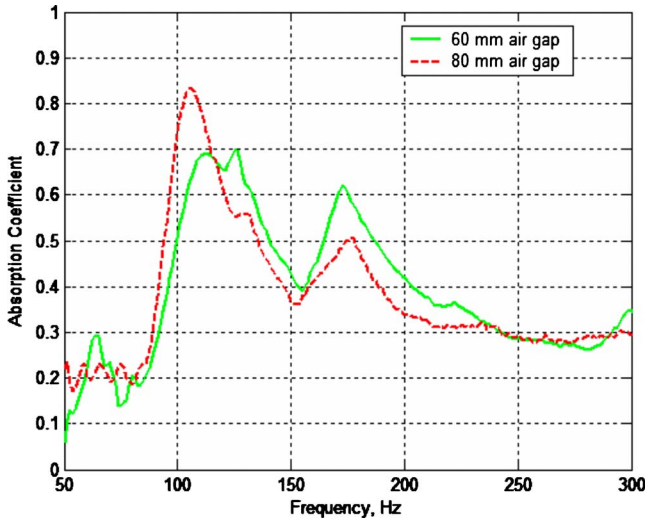


FIG. 7. (Color online) Measured absorption coefficient spectra corresponding to Fig. 6 (Ref. 7).

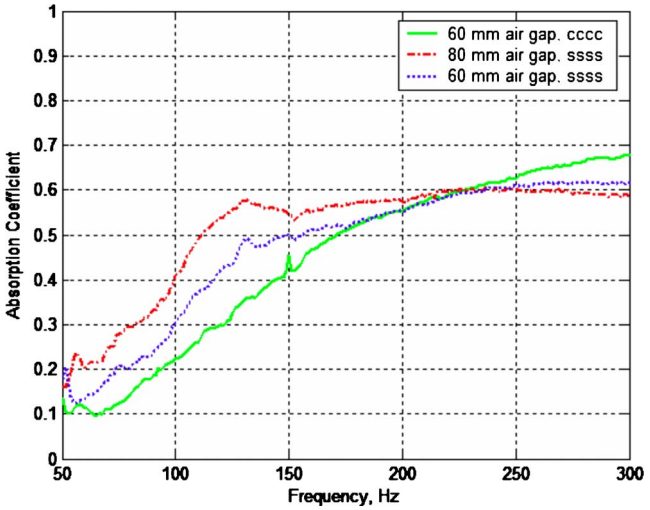


FIG. 8. (Color online) Measured absorption coefficient spectra obtained with a 60 cm square Black plate with 60 and 80 mm wide air gaps and rigid (CCCC) and simply supported (SSSS) edge conditions.

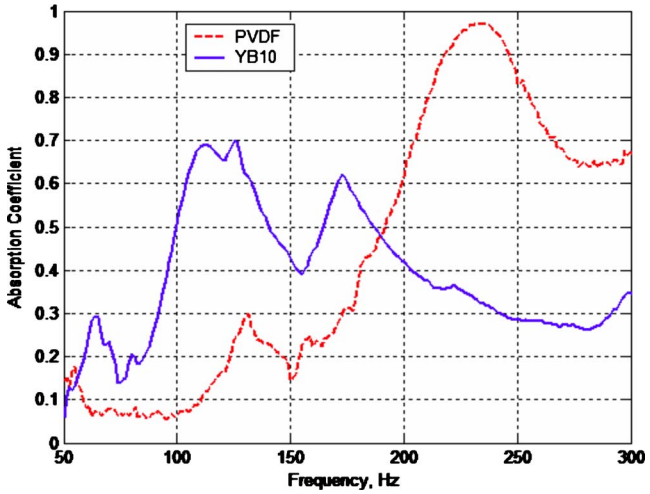


FIG. 10. (Color online) Measured absorption coefficient spectra corresponding to Fig. 9 (Ref. 7).

than those for the PVDF foam plate. The PVDF plate yields a lower absorption coefficient than the YB10 plate below 190 Hz. However, the major resonance in the absorption coefficient of the PVDF plate at 232.5 Hz means that it has a higher absorption coefficient than the YB10 foam plate above 190 Hz. There are minima in the PVDF plate absorption spectrum at 150 and 280 Hz compared with 70 and 155 Hz for the YB10 plate.

V. COMPARISONS BETWEEN PREDICTIONS AND DATA

Figure 11 compares the measured absorption coefficient spectra of the rigidly clamped YB10 plate with those predicted for fluid loaded, rigidly clamped, rectangular porous elastic plate using the properties listed for the YB10 plate in Table I. Also shown are predictions for a hypothetical nonporous plate having the same elastic properties as the porous elastic plate used in the measurements. Although the agreement between predictions and measurements is not particularly good, the predictions that include the effects of porosity are clearly superior. They match the frequency and magnitude of the peak between 100 and 150 Hz fairly well and the relative magnitudes of the other two peaks. The predicted absorption coefficient for a hypothetical nonporous plate having the same elastic properties as the porous elastic plate is below 0.15 throughout the frequency range. This suggests that the presence of porosity increases the absorption of the poroelastic plate configuration. The depth of the air gap used when obtaining the data shown in Figs. 11 and 12 was 60 cm.

Figure 12 compares the measured absorption coefficient spectra of the PVDF plate, under intermediate clamping, with those predicted for fluid loaded, rigidly clamped, rectangular porous and nonporous elastic plates having the properties listed for PVDF plate in Table I. Again the predictions including porosity are superior in that the frequency and

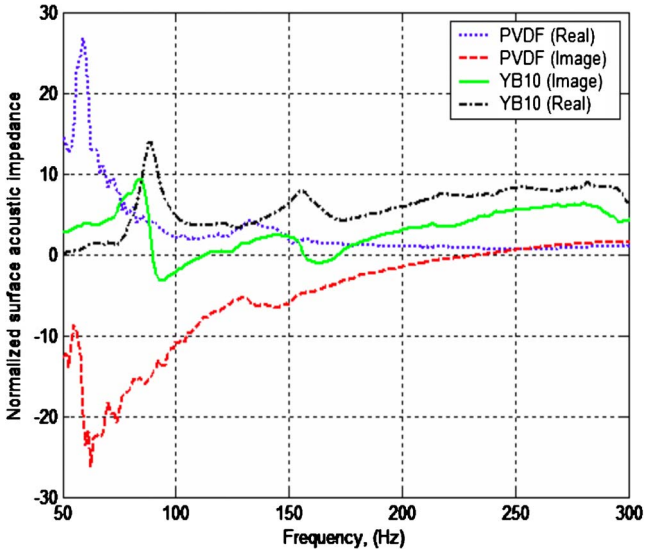


FIG. 9. (Color online) A comparison of measured normalized surface acoustic impedances of the PVDF and YB10 foam plates with 60 mm wide air gap behind the plates and intermediate clamping conditions.

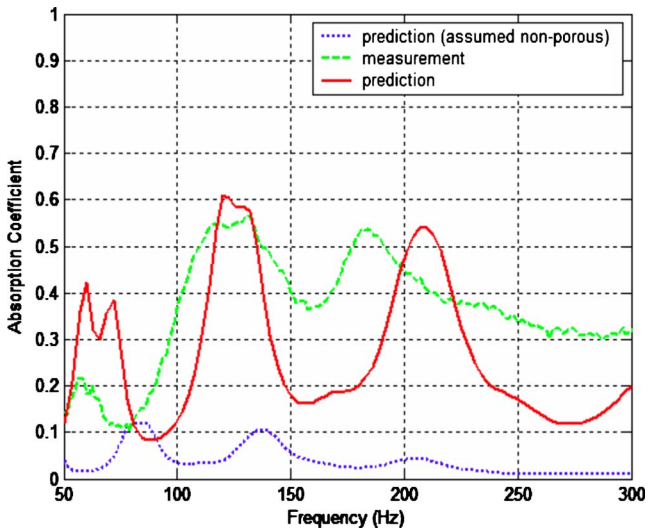


FIG. 11. (Color online) A comparison of the measured and predicted absorption coefficient spectra for a rigidly clamped YB10 plate. Predictions are shown for both porous and nonporous plates.

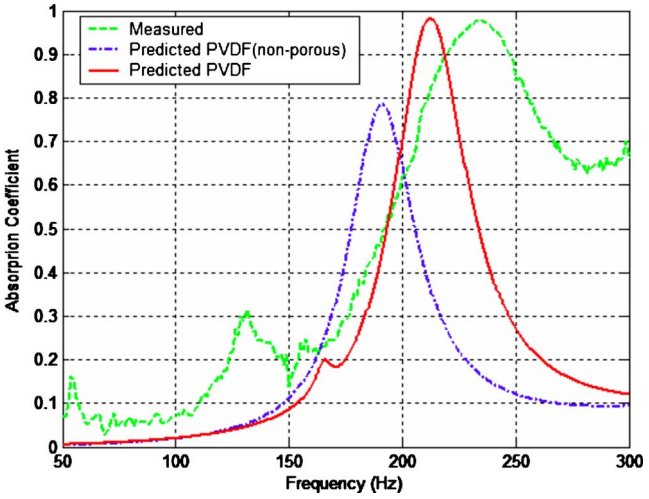


FIG. 12. (Color online) A comparison of the measured absorption coefficient of the PVDF plate, under intermediate clamping conditions, to the predicted absorption coefficients of rigidly clamped porous and nonporous PVDF plates.

magnitude of the resonant peak are closer to those measured. However, the predicted influence of porosity is much less for this predominantly closed cell foam. An additional reason for differences between the predictions and data is that the theory assumes open cells, whereas the PVDF plate has predominantly closed cells.

To investigate the extent to which the surface impedance resonances are related to plate vibration, measurements have been made of the deflection spectra while the plates were clamped in the impedance tube. Figure 13 shows the deflection at the center of the YB10 plate measured using an accelerometer while it was clamped in the impedance tube and subject to broadband sound. Also shown is the deflection spectrum predicted by using Leclaire's theory[5] modified to include fluid loading[6,7] and assuming a uniform oscillatory driving pressure corresponding to a plane wave at normal incidence. The observed and predicted deflection maxima at 60, 120, and 170 Hz are close to the frequencies (58.75, 130,

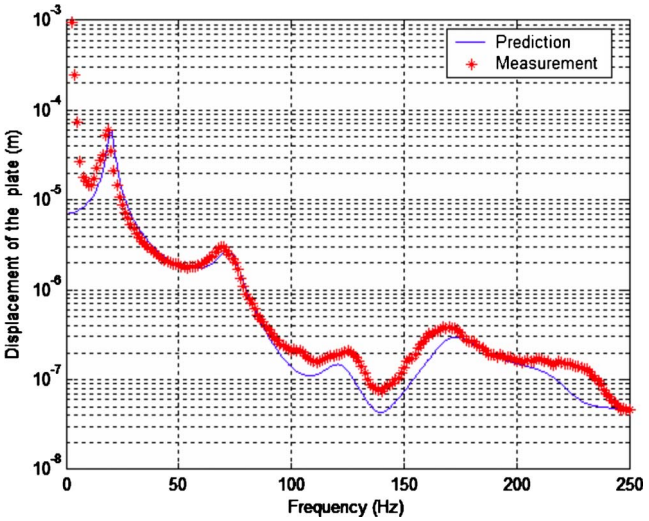


FIG. 13. (Color online) A comparison of measured and predicted deflection spectra of a rigidly clamped YB10 plate subject to uniform oscillatory pressure (Ref. 8).

and 170 Hz) of the peaks observed in the absorption coefficient spectra for the YB10 foam plate (Fig. 7).

To further emphasize the importance of the elasticity to the acoustical performance of the clamped porous plates, calculations have been made also of the surface impedance and normal incidence absorption coefficient of a rigid-porous plate with the same pore-related properties as the porous elastic plate. Figure 14 shows predictions, using the Johnson–Allard model,[9] of the surface impedance and absorption coefficient spectra below 300 Hz of a rigid-porous plate with and without a 60 mm wide air gap in front of a rigid termination. The plate is assumed to have the same pore-related properties as the YB10 plate (from Table I). Figure 14 shows that, even with an air gap, the absorption coefficient for the rigid-porous plate is predicted to be below 0.45 throughout the frequency range. The fact that these predictions do not contain any peaks provides further confirmation that the peaks observed in measurements (Figs. 4–12) are related to plate vibration resonances and that to obtain useful low frequency absorption peaks a clamped plate should be porous, elastic, and have a backing air gap.

The low frequency absorption peaks obtained with the Black plate are significantly less pronounced than those obtained with the YB10 plate even with the same boundary conditions. There are differences in stiffness, thickness, and porosity-related parameters. The YB10 plate is much stiffer than the Black plate. Using Eqs. (1)–(4), the predicted effect of decreasing Young's modulus only is to increase the absorption coefficient peaks and move them to higher frequency. The thickness of the Black plate is almost double of the YB10 plate. The predicted effect of only increasing the thickness of the YB10 plate from 10.7 to 20 mm is to decrease the absorption coefficient of the YB10 plate to below 0.1 throughout the frequency range. Porosity has an important effect on predictions of the absorption coefficient only if it is varied independent of the flow resistivity. However, this is unphysical since flow resistivity will depend on porosity as well as on cell structure. Change in flow resistivity only is predicted to have a relatively small effect on the absorption coefficient. On this basis, the observed differences in absorption spectra between Black and YB10 plates can be attributed primarily to differences in thickness and stiffness.

The PVDF and YB10 plates have different Young's modulus, thickness, porosity, density, flow resistivity, and cell structure. Specifically, the PVDF plate has higher flow resistivity and thickness but lower porosity and Young's modulus. For a given (intermediate) clamping condition, although the predicted effect of decreasing Young's modulus alone is to increase the absorption coefficient peaks and move them to higher frequency, the predicted effect of increasing thickness alone is to reduce the absorption peaks, as observed when comparing the YB10 plate performance with the Black plate. Moreover the predicted effect of increasing flow resistivity is not very significant. The comparative failure of theory, which assumes an open cell structure, to predict the acoustical performance of the clamped PVDF plate seems, therefore, to be related to its predominantly closed cell structure.

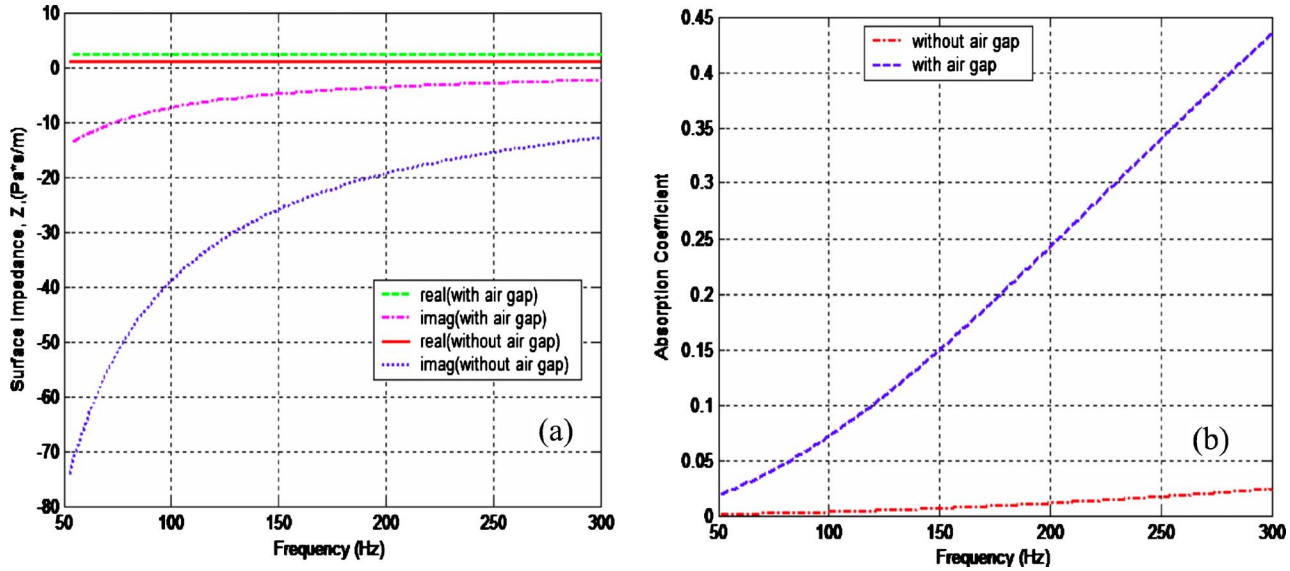


FIG. 14. (Color online) Predicted (Ref. 9) surface impedance and (b) absorption coefficient spectra of a rigid-porous plate having the pore-related properties of the YB10 plate with and without an air gap.

VI. CONCLUSIONS

Measurements have been made of the absorption coefficient and the surface acoustic impedance of clamped porous elastic plates with an air gap in front of the rigid termination of a large impedance tube. These show resonances below 300 Hz that may be useful for noise control and are similar to those previously reported.[1] Here they have been shown to be at frequencies close to those of the maxima measured and predicted in the plate deflection spectra. A hybrid model, formulated by incorporating the predictions of clamped fluid loaded poroelastic plate vibration in the expression for surface impedance of an equivalent fluid plate of infinite length, has been found to give tolerable agreement with data. The discrepancies between predictions and measurements may be due to several factors including assumptions made in the theory, uncertainties in plate properties, and poor practical realizations of the assumed boundary conditions. However, the relative insensitivity of the predictions to plate parameters suggests that the other two causes are likely to be the more important ones. The low frequency resonances have been shown to depend on the clamping conditions and the width of the backing air gap in addition to the pore- and elasticity-related properties and thickness of the plates. Increasing the rigidity of the clamping is found to reduce the low frequency absorption peaks. The dependence on clamping conditions may explain the fact that, although two of the materials tested (YB10 and Black) are nominally similar to those tested previously by others,[1] contrasting results have been obtained here. Three low frequency absorption peaks have been observed for the YB10 plate compared with the single relatively small low frequency absorption peak observed elsewhere.[1] Moreover, according to our data, Black plates perform less well than YB10 plates. An intermediately clamped 25 mm thick plate of closed cell foam (PVDF) has been found to give a substantial absorption coefficient peak below 300 Hz. The markedly different absorption coefficient spectra observed with the predominantly closed cell PVDF plate compared with the YB10 plate for similar clamping conditions appear to result from the different cell structures rather than from differences in stiffness, flow resistivity, and thickness and this should be the subject of further work.

ACKNOWLEDGMENTS

This work was supported in part by EPSRC (UK) Grant No. GR/R43761. The authors are grateful also for an RAIS award supported by EPSRC (UK) and Armacell (UK).

[1]M. J. Swift, K. V. Horoshenkov, P. Leclaire, D. C. Hothersall, K. Fujiwara, and H. Torihama, "On the effect of the bending vibration on the acoustic properties of thin poroelastic plates," J. Acoust. Soc. Am. **107**, 1786–1789 (2000).
[2]D. Takahashi and M. Tanaka, "Flexural vibration of perforated plates and porous elastic materials under acoustic loading," J. Acoust. Soc. Am. **112**, 1456–1464 (2002).
[3]K. V. Horoshenkov and K. Sakagami, "A method to calculate the acoustic response of a thin, baffled, simply supported poroelastic plate," J. Acoust. Soc. Am. **110**, 904–917 (2001).
[4]K. V. Horoshenkov, K. Sakagami, and M. Morimoto, "On the dissipation of acoustic energy in a thin, infinite, poroelastic plate," Acta. Acust. Acust. **88**, 500–506 (2002).
[5]P. Leclaire, K. V. Horoshenkov, M. J. Swift, and D. C. Hothersall, "The vibrational response of a clamped rectangular porous plate," J. Sound Vib. **247**, 19—31 (2001).
[6]H. Aygun, K. Attenborough, and A. Cummings, "Predicted effects of fluid loading on the vibration of elastic porous plates," Acta. Acust. Acust., **93**, 284–289 (2007).
[7]H. Aygun, "The design of noise attenuating devices incorporating elastic porous structures," Ph.D. thesis, University of Hull, Hull, UK (2006).
[8]K. Attenborough and H. Aygun, "Acoustical performance of clamped and perforated poroelastic plates," 19th International Congress on Acoustics, Madrid, September 2–7 2007. http://www.sea-acustica.es/WEB_ICA_07/fchrs/papers/rba-09–001.pdf. Last accessed 6/27/08.
[9]J.-F. Allard, *Sound Propagation in Porous Media* (Elsevier Applied Science, New York, 1993), pp. 76–93 and Eqs. (5.62)–(5.66).
[10]Designation: E 1050-90, "Standard test method for impedance and absorption of acoustical materials using a tube, two microphones, and a digital frequency analysis system."

On the acoustics of ancient Greek and Roman theaters

Andrea Farnetani,[a] Nicola Prodi, and Roberto Pompoli
Dipartimento di Ingegneria, Università degli Studi di Ferrara, via Saragat, 1-44100 Ferrara, Italy

(Received 22 May 2007; revised 4 June 2008; accepted 4 June 2008)

The interplay of architecture and acoustics is remarkable in ancient Greek and Roman theaters. Frequently they are nowadays lively performance spaces and the knowledge of the sound field inside them is still an issue of relevant importance. Even if the transition from Greek to Roman theaters can be described with a great architectural detail, a comprehensive and objective approach to the two types of spaces from the acoustical point of view is available at present only as a computer model study [P. Chourmouziadou and J. Kang, "Acoustic evolution of ancient Greek and Roman theaters," Appl. Acoust. **69**, re (2007)]. This work addresses the same topic from the experimental point of view, and its aim is to provide a basis to the acoustical evolution from Greek to Roman theater design. First, by means of *in situ* and scale model measurements, the most important features of the sound field in ancient theaters are clarified and discussed. Then it has been possible to match quantitatively the role of some remarkable architectural design variables with acoustics, and it is seen how this criterion can be used effectively to define different groups of ancient theaters. Finally some more specific wave phenomena are addressed and discussed.
 [DOI: 10.1121/1.2951604]

PACS number(s): 43.55.Gx, 43.55.Ka, 43.55.Mc [NX] Pages: 1557–1567

I. INTRODUCTION

Greek and Roman theaters had a role of paramount importance in the history of western culture. Many of such spaces are still used nowadays for their original scope with minor or no adaptations at all. The outstanding achievement of ancient designers to provide such robust and outstanding spaces of performance has long attracted architects and, in this framework, acoustics plays a key role. In fact, both for their architecture and its related acoustics, the Greek theaters can be considered the starting point of theater history and design.[1] The peculiar acoustical character of the ancient theaters has been the subject of several specific studies in the past. One of the most important theoretical works in this respect is Canac's book[2] where the geometrical room acoustics approach was applied. In this study a basic energetic equation for the sound level was developed as a function of the hearing angle and of the slope of the *cavea*. Sometimes the design skill inherited from the past was valorized. This is the case, for instance, for the design of the cavea and of the Greek technique of increasing its slope toward the farthest positions, which is regarded as a valid rule also for modern theater design.[3] Furthermore one can find in literature several reports on acoustical measurements and on computer simulations applied to ancient theaters. In Ref. 4, acoustical measurements in the Theater of Epidauros are presented. Sound level, speech intelligibility, and reverberation time are analyzed, and the conclusion is that the reason for good acoustics lays on the achievement of Greek designers to optimize theaters plan and section. Then Shankland,[5] on the basis of speech articulation score measurements, hypothesized that the early scattering from the edge steps plays an important role. More recently, measurements in the Aspendos, Perge, and Termessos Theaters were presented in Ref. 6 together with acoustical simulations based on computer models. Measurements of orthogonal factors were done in Delphi and Taormina.[7] Other works used computer simulation to study the sound field in ancient theaters as Ref. 8 where the acoustical properties of the Epidauros theater are analyzed and compared to two other theaters (Dodoni and Patras Odeon) via computer-aided prediction and auralization. Finally, a 1:10 scale model of the Delphi theater with 5000 scale-sized seated figures was built.[9] Furthermore a specific issue of the sound field in ancient theaters was recently addressed, that is, the effect of diffraction of the seat rows.[10] Although those effects are not specifically studied in the present paper, they will show up in the obtained results and will be discussed. All in all the review of literature provides a useful outlook on the main features of either Greek or Roman theaters, but rarely those two types of spaces have been compared directly in a single work. A notable exception is the recent paper[11] where the acoustical evolution from the former to the latter was investigated by the use of computer modeling. After the use of such tool was formerly assessed, several models of ideal ancient theaters in their original configuration were developed and systematically studied. The study[11] makes very limited use of acoustical measurements and refers to ideal ancient theaters, complete in every part. Actually such theaters are hardly found today due to the consequences of time degradation or to successive refurbishments with different styles.

The aim of the present work is to investigate the change of the acoustics from Greek to Roman ancient theaters and, in particular, to trace back the acoustical characteristics to specific design elements introduced during this architectural evolution. The work is based on *in situ* acoustical measurements in a group of five Greek and Roman ancient theaters and on scale model measurements. In fact, thanks to a rich

[a]Electronic mail: andrea.farnetani@unife.it

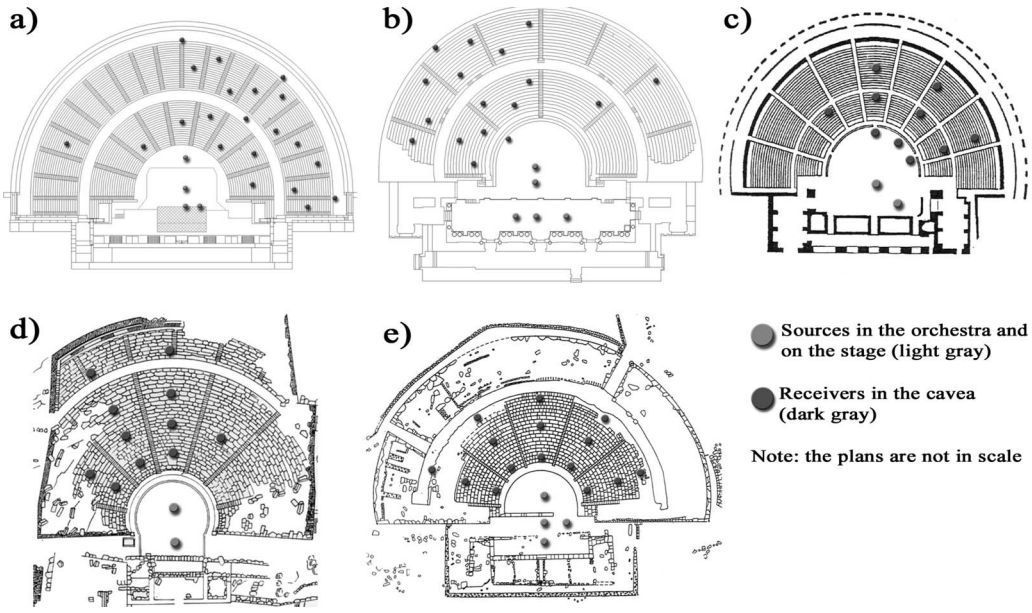


FIG. 1. Plans of selected ancient theaters: (a) Aspendos, (b) Jerash, (c) Taormina, (d) Delphi, and (e) Segesta.

archaeological documentation of the ancient theater of Siracusa (Italy),[12] it was possible to focus on the architectural evolution of this theater across the centuries from Greek to Roman times. Based on the archeological evidences quite specific historical configurations of the theater of Siracusa were identified and a scale model was built to test them acoustically. Unfortunately this theater is today lacking many original parts and *in situ* surveys would not be representative of historical state. The scale model measurements of the different layouts of the Siracusa theater and the *in situ* measurements in the five real theaters were collected into one data set. By doing so the outstanding role of the several architectural elements was isolated. Part of the present work was developed within the EU Project ERATO (Identification Evaluation and Revival of the Acoustical Heritage of Ancient Theaters and Odea)[13] whose aim was the acoustical and visual reconstruction, in a three dimensional virtual environment, of some outstanding examples of ancient theaters and odea.

II. *IN SITU* MEASUREMENTS

Within the ERATO project, two well preserved theaters were chosen for acoustical measurements. The former is located in Aspendos in the southern part of Turkey and the latter is the Jerash South Theatre in Jordan. In particular, the theater of Aspendos is probably the best preserved Roman theater in the world as it is the only one having a complete original stage wall and also an aisle covered by a colonnade in the upper part of the cavea.

Moreover, acoustical measurements were taken in three other ancient theaters: the Greek theater in Delphi (Greece), the ancient theater in Taormina, and the ancient theater in Segesta, both located in the South of Italy. In this work those measurements are grouped with the results from the ERATO project to have a richer set of data. The plans of the theaters measured *in situ* are shown (not in scale) in Fig. 1. It can be noted that the layout is semicircular for all the theaters except for Delphi, which shows a lightly pronounced fan shaped cavea. The main architectural characteristics of the theaters are reported in Table III. The following acronyms will be used to identify the different theatres: Aspendos (AS), Jerash (JE), Taormina (TA), Delphi (DE), and Segesta (SE).

The measuring equipment used during the sessions consisted of a portable computer and a multichannel sound card. Moreover a dodecahedron sound source and two different types of microphones were used. In fact, acoustical data were acquired in the binaural and monaural formats as well as in the B-format one, to have a complete characterization of the sound field and, in addition, to achieve useful primary data for auralization.

A 20 s long exponential sweep sine signal was generated from 50 Hz to 20 kHz and sent to the source amplifier. Then the four channels of a B-format microphone and the two channels of a dummy head were simultaneously recorded and stored for postprocessing. For each source-receiver position the impulse response was obtained by convolving the acquired signal with the suitable inverse filter. Finally several acoustical parameters were calculated based on the monaural W signal of the B-format impulse response according to the ISO3382-1 standard.[14]

A different number of sources and receivers were used depending on theaters size. Two source positions are common to all measured theaters: the former on the front stage and the latter in the central area of the orchestra. The height of the source was 1.5 m. The microphone probes were placed at 0.7 m from each other and all the signals were recorded simultaneously. Moreover the probes were 1.1 m high above their reference floor (the step where the listeners' feet rest). All the theaters were unoccupied during the measurements.

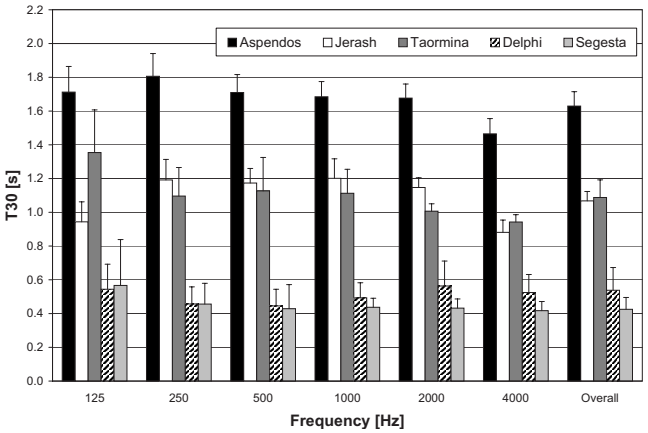


FIG. 2. Comparison of the average reverberation time for all theaters.

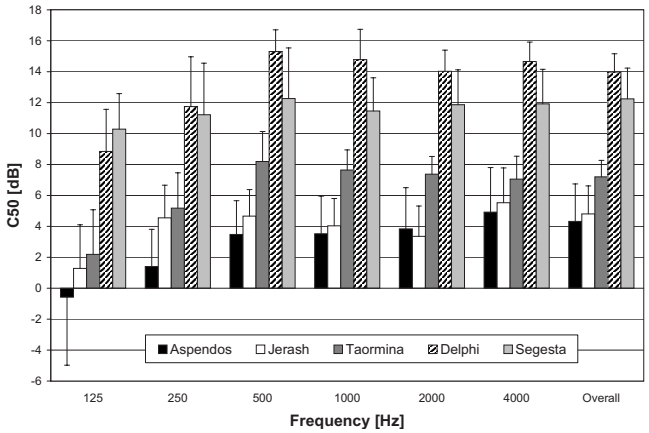


FIG. 3. Comparison of the clarity parameter for all theaters.

III. RESULTS FROM *IN SITU* MEASUREMENTS

A. Reverberation

The values of reverberation time T_{30}, as defined in Ref. 14, obtained averaging all sources and receivers, are reported in Fig. 2 as a function of frequency in the octave bands from 125 Hz to 4 kHz (the overall value is also reported). The standard deviation bars of the spatial distribution of the parameters are included. The bars extend from the top of each column. AS has the higher T_{30} (about 1.6 s at middle frequencies) because it is the only one that has a complete stage building and the upper colonnade. The theater is laterally closed and the sound energy can circulate inside, undergoing many scattered reflections that cause a longer reverberant tail. The T_{30} of JE and TA is almost coinciding (1.2 s at middle frequencies) even if their dimensions are different. In fact, the cavea in TA is nearly as wide as in AS (inner diameter about 100 m) whereas in JE it is only 63 m. Nevertheless their overall architectural features in the present state are similar: both have a partial stage wall and have no upper gallery.

Finally, the lower T_{30} (about 0.5 s for middle frequencies) was found in the DE and SE as they have no stage building. The sound scattered from the tiers of steps does not undergo further reflections and the energy is rapidly lost.

B. Clarity

The C_{50} parameter, as defined in Ref. 14, is preferred in this study as the clarity for speech is more suitable in ancient theaters, where spoken word had a special importance. In Fig. 3 the average values of the C_{50} calculated for all source and receiver positions are presented as a function of frequency. The spatial standard deviation is also shown for each frequency band. First it can be seen that in AS, although the reverberation time is rather high, the values of C_{50} are also remarkably high. This fact is due to the most energetic reflections which all arrive with limited delay in the very first part of the impulse response. Moreover the graphs show a great variability of the parameter regarding the receiver positions, with a standard deviation of about 2 dB for all the theaters. Very high clarity values are found in DE and SE (about 14 and 12 dB for middle frequencies, respectively) in accordance with very low measured T_{30}. Finally, JE and TA have different clarity (4 dB higher in the second theater) although they had nearly the same T_{30}. This could be probably explained considering the different height of the stage wall which is more than double in TA compared to JE.

C. Sound strength

Another interesting analysis can be carried out by considering the sound propagation curves in the five theaters. For this scope the parameter G_{mid}, which is the average of G, as defined in Ref. 14, for the middle frequencies (500 Hz, 1 kHz, and 2 kHz octave bands), is calculated. All the data, measured for each source-receiver combination, are shown in Fig. 4. For ease of comparison in Fig. 5 only the regression lines obtained with a linear fit of the measured data are reported. In Table I the values of the parameters A_i and B_i of the equation $G_{\mathrm{mid}}=A_i+B_i \log_{10}(\mathrm{dist})$ are reported for each theater together with the related regression coefficients. Despite the different measured T_{30}, the energy levels are fairly similar and are comprised within a range of 4 dB. This prompts the idea that the sound energy is mainly concentrated on the first part of the impulse response, including the direct sound and the two outstanding reflections from the

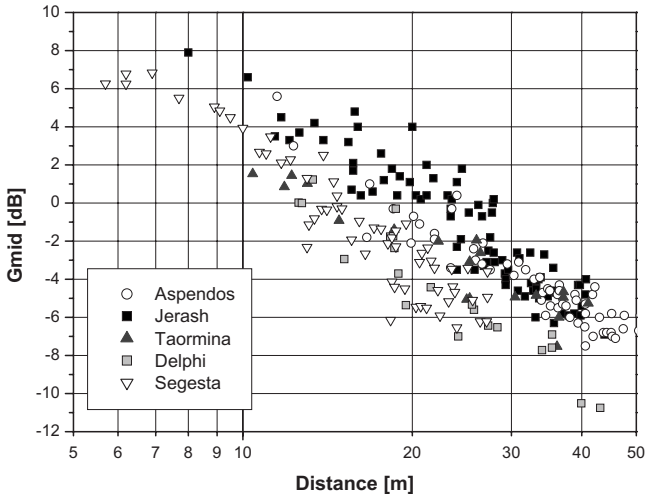


FIG. 4. The sound strength measured at different positions in all theaters at middle frequencies.

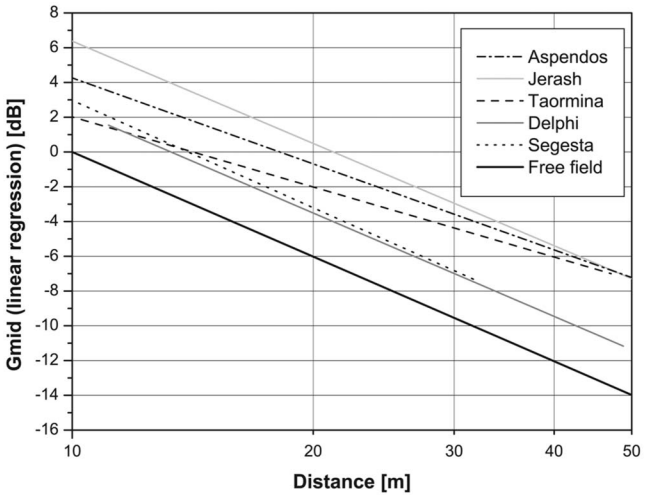


FIG. 5. Comparison of the sound strength regression lines for all theaters at middle frequencies.

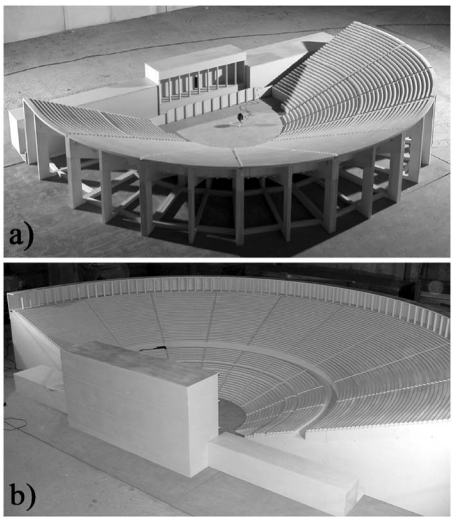


FIG. 6. Examples of modeled theaters configuration: (a) Greek and (b) Roman.

floor and the stage building (when present), and only a smaller amount of energy comes from the tail. For this reason the lower levels were registered in DE and SE because of the lack of the reflection from the stage wall. The G_{mid} propagation curve in JE has actually the same slope as in DE but it is about 4 dB higher. Moreover, for the shortest distances the sound level in JE is the highest. In fact, having this theater the highest cavea slope in the group, for the same source-receiver distance the stage wall will be closer to the audience and the reflection from this surface will have more energy at the receiver positions. On the contrary, for the farthest distances, in the AS theater the reflections coming from the colonnaded gallery cause the level rise and make the propagation curve less steep.

IV. SCALE MODELED THEATERS

The ancient theater of Siracusa, situated in the south of Italy, is one of the biggest theaters of the *Magna Grecia*. The theater had undergone major changes and fortunately the site excavations made it possible to trace with great detail three different layouts dating from the fifth century B.C. to the second century A.D.,[12] that is a Greek, a Hellenistic, and a Roman theater. Consequently, a modular scale model of the Siracusa theater was built to deeply investigate in a single site the acoustical changes from the former Greek to the latter Roman design.

The Siracusa Greek theater (SG), also called Timoleonte theater, is dated about 335 B.C. It had 37 tiers of steps with a slope of 20°. The maximum diameter of the semicircular cavea was about 85 m and its height was 11 m at the last step. The radius of the orchestra was approximately 12 m. The rows of steps were continuous, without any corridor (*diazoma*), and the entrances (*parodoi*) to the cavea were from the upper part of the hill and on the side of the stage building. This was in turn 23 m wide and 8.6 m high, with a row of ten columns on the front supporting a portico 3 m deep. In front of it, at a distance of 3.5 m, there was the wooden parapet (2.5 m high) which supported a wooden floor during the public meetings and hid a wagon, having the function of a moving support for the set, during the venues. It can be estimated that the SG had space for about 5000. A view of the model representing SG is shown in Fig. 6(a).

TABLE I. Linear fit parameters for measured theater sound level decays $[G_{\mathrm{mid}}=A_i+B_i \log_{10}(\mathrm{dist})\ (\mathrm{dB})]$.

Theater	A_i (dB)	B_i (dB)	R
Free field	20	20	1
Aspendos	20.68	−16.43	−0.93
Jerash	25.93	−19.55	−0.92
Taormina	15.41	−13.99	−0.93
Segesta	23.49	−20.52	−0.94
Delphi	22.19	−19.75	−0.93

The Siracusa Hellenistic theater (SH), also called Gerone II theater, was built from 238 to 214 B.C. basically as an expansion of SG. Some 31 tiers of steps were added to the semicircular cavea which summed up to 68. Thus the capacity was greatly increased up to nearly 15 000. Despite its Greek origins, the Hellenistic theater had a covered gallery with a colonnade in the upper part of the cavea, as usually found in later Roman theaters. This feature makes the SH theater peculiar with respect to other Hellenistic theaters. As shown later this will imply a closer acoustical similarity with Roman theaters. A proper diazoma, called *diazoma major*, was also added by removing the step rows from 29 to 31 and two additional entrances (parodoi) were created at the lateral ends of it. The slope of the cavea was kept at 20° and no changes were made in the orchestra. Now the maximum diameter of the cavea was 143 m and it was 26 m high, including the columns (whose height was about 5 m). The stage building was enlarged up to the lateral *piloni* (the two rocky bodies flanking the scenic area). The parodoi, as vaulted corridors, became internal to the building which grew up to 28 m wide and 15.4 m high. A stage set 3 m high and 3 m deep was built with stones. Because of the few remains, it is difficult to know exactly how the facade looked like, but probably the skene had a monumental appearance.

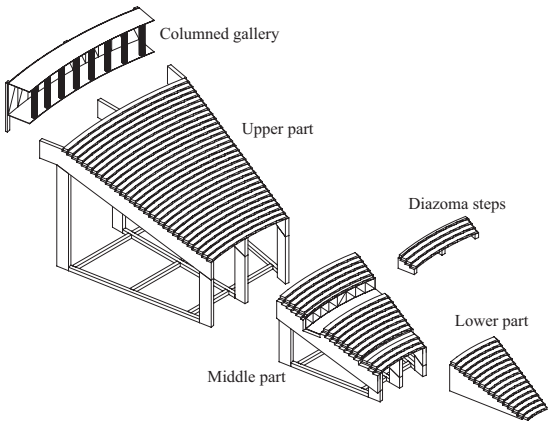


FIG. 7. The structure of a "slice" of the model.

In the second century A.D. the theater was again modified to meet the Roman style. Due to an increased need of space in the orchestra, the first four seat rows were removed. In order to join the cavea with the orchestra its slope changed from 20° to 25°. So the Siracusa Roman theater (SR) had 64 tiers of steps and the radius of the orchestra was 15.3 m. Furthermore, row 20 was removed to create the *diazoma minor*, which was nine tiers of steps under the diazoma major. Two tunnels (the *cryptae*) were carved on the sides of the stage building, under the cavea seats, to create two new parodoi. The stage building was again enlarged. It had a basement with a stage front and three storeys with columns on the facade. It was 41 m wide and about 20 m high. The stage set was 1.5 m high and 7 m deep. The number of seats was almost unchanged. Figure 6(b) shows the model of SR: it is not archeologically proved that the stage building was connected to the cavea for the full height.

V. SCALE MODEL ASSEMBLY AND MEASUREMENT SETUP

A. The modular structure

In order to reproduce the historical configurations described above and derived from the archaeological evidences the scale model was designed as a modular structure. This means that by adding or removing some structural parts such as tiers of steps, the colonnaded gallery, and different stage buildings, it is possible to assemble each of the above theaters. Moreover also some hybrid configurations, which were not historically proved, could be assembled to investigate the acoustical contribution of specific architectural elements.

A 1:20 scale was chosen as the best possible compromise between the contrasting requirements of the space to settle the model on the one hand and the measurable frequency range on the other hand.

Basically, the cavea of the model consists of nine different circular sectors having an angle of 20.5°, each divided into three main parts (Fig. 7): the lower, the middle, and the upper tiers of steps. Actually two different versions of the lower parts were prepared in order to take into account the different slopes of the steps introduced in SR. Moreover, considering that the lateral parts of the cavea do not have a perfectly circular shape but end with a straight zone, the two lateral slices are different from the others. Movable parts were also designed to be able to modify the cavea structure and to introduce the two diazoma added in Hellenistic and Roman theaters, respectively. As it is visible in Fig. 7, these parts consist in a set of three and a half steps for the diazoma major and one step only for the diazoma minor. Each of the three parts, composing all the steps in a circular sector, are supported by a structure made of beams.

Starting from SG, its cavea is built by joining, for each slice, the lower and the middle part and adding then together the nine slices. Then, for achieving SH, the upper part of the cavea and the columned gallery are added and three rows of steps are removed to create the diazoma major. Finally, the SR is obtained by swapping the first tiers of steps and creating the diazoma minor by removing the 20th step. For each configuration the appropriate stage building was added in front of the respective cavea. The three different configurations of the scale model are shown in Fig. 6. The structure of wooden beams supporting the cavea and the split in nine slices can also be observed in the picture of SG.

B. Scale model approach for open theaters

In order to take acoustical measurements in a scale model, it is necessary to use a measurement chain for high frequency. This can be deduced from a few geometrical and acoustical rules, which define the transformation of full-size quantities to model quantities.[15] Moreover, the sound absorption in the propagation medium has to be taken into account and two methods have usually been applied.[16] Unfortunately, the model of an ancient theater is not a closed room and furthermore it needs a very big space to be stored. Thus, it was not possible to dry the air or to fill with nitrogen such a big volume. Another possible method to correct for the air absorption is the numerical compensation of the acquired impulse responses.[17] This method was first experienced but later dropped. Its effectiveness in the present case is, in fact, not as important as in closed spaces. The reason for this is that the absorption of the missing ceiling is far larger than any possible correction for the air absorption. This scale model approach for open air theaters was validated by means of both computer simulations[18] and real scale measurements. Unfortunately the Siracusa theater today does not correspond to any configuration of the scale model, so that a direct acoustical comparison of the real theater with the scale model is not possible. Anyway the validation stems from the comparison of the measurement results in the scale model with the data obtained for the respective typology of real theaters analyzed in this work. In fact, the results showed a very good consistency up to a frequency limit set in the 2 kHz band (real scale). From this band on the discrepancies due to the air absorption started to play a not negligible role. Finally, the environmental conditions were very stable during the measurement sessions and this improved the reliability of the results.

C. Scale model measurement setup

Two different types of sound sources were used in the measurements. The first one was a spark generator whose

spinterometer had an omnidirectional emission on a horizontal plane. Moreover, the discharge height was 75 mm high corresponding to 1.5 m in the real scale, which is the source height adopted for the measurements in the real theaters. The frequency response, not reported, is not flat, but it rises by about 9 dB per octave band, almost compensating the air absorption at high frequencies. Unfortunately the repeatability of the sparks is not good (level differences within 20%) and this source cannot be employed to measure the sound levels in the scale model. So this device was primarily used to analyze the fine structure of the impulse responses because the spark is very sharp and clean making it possible to investigate in detail the pattern of reflections.

The second source was a miniaturized dodecahedron, having a diameter of about 70 mm, made of piezoelectric transducers covered by a layer of paper. The frequency response of the source, measured at 1 m, is rather flat from 2.5 to 50 kHz, which corresponds to 125 Hz–2.5 kHz in the real scale, considering the scale factor of the model.

For each theater configuration two source positions were chosen. The former is in the orchestra, close but not coinciding with the geometrical center, to avoid the focusing effects caused by the semicircular shape. This position is exactly the same for all the theaters to allow direct comparisons. The latter position was on the stage in the SH and SR. Since SG has no front stage (the area is occupied by the system for moving the stage set), the latter position was located in the rear part of the orchestra close to the stage. The height from the floor to the acoustical center of the source was 7.5 cm too.

An eight channel sound card with 192 kHz sampling frequency was used since this frequency is the highest available within the professional audio standards nowadays. Thus, having a sample rate of 192 kHz, the higher frequency becomes 96 kHz. This means that passing from the scale model to the real dimensions the higher working frequency is 4.8 kHz. Considering the better combination of size, sensitivity, and frequency range, also in relation to the sound sources described above, a 1/4 in. condenser microphone was chosen for measurements in the scale model. The microphone was pointed toward the source and in each measurement position it was placed two steps above so that the diaphragm was exactly in the position occupied by the head of a virtual listener sitting on that step.

The measurement positions were distributed in the cavea covering all possible listening positions and were regularly spaced within each slice of the theater. In SG the measurement positions were 36 and in SH 15 more were added for a total of 51. In the SR, where the first rows changed, the number of receivers was 42.

VI. RESULTS FROM SCALE MODEL MEASUREMENTS

In this section the results of the measurements in the unoccupied scale model will be presented. First the model was configured as SG, SH, and SR and then some hybrid configurations were assembled (i.e., removing or adding, for example, the stage building or the upper gallery) and acoustically tested. The data are translated into the full scale frequencies and distances.

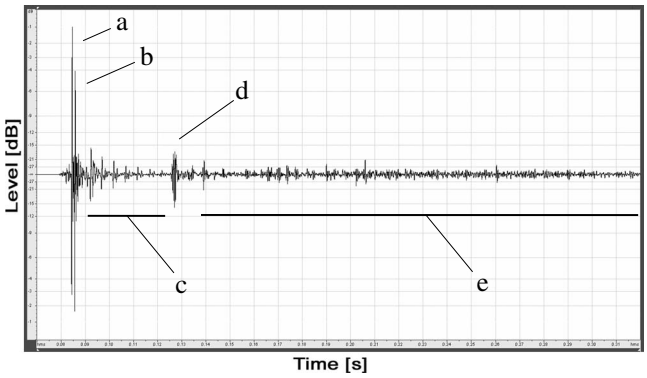


FIG. 8. Example of IR from SR: source in the orchestra, receiver on the fifth step (first 250 ms). (a) Direct sound, (b) orchestra floor, (c) upper steps, (d) stage building, and (e) reverberant tail.

First of all the reflection pattern of a typical impulse response measured in the modeled theaters has to be considered. An example of the first 250 ms of an IR recorded in SR (source in the orchestra, receiver on the fifth step) is shown in Fig. 8. From the analysis of the pattern of reflections it appears that after the direct sound (a) there are only two major reflections: the former comes from the orchestra floor (b) and the latter, less loud and delayed too, is from the stage building (d). Some reflections are also recognized between the floor and the stage building reflections (c). A geometrical verification of the delay times of these reflections showed that the peaks correspond exactly to seven step edges behind the microphone position. The reflection of the closer seat riser and those of the steps in front of the microphone are completely masked by the direct sound and by the floor reflection. After the stage building (d), in the impulse response one can recognize a group of distributed reflections with much smaller amplitude (e). They make up the reverberant tail and are mainly caused by the multiple scattered reflections of sound on the tiers of steps in the cavea. Keeping this figure in mind, the results of the measured acoustical parameters can be presented.

A. Reverberation

For the scale model the reverberation time T_{20} is used as in some cases there was not enough dynamics to measure T_{30} in a reliable way. A comparison of T_{20} measured in SG, SH, and SR is shown in Fig. 9. The data are the average of the two sources and all the receivers and the bars correspond to the spatial standard deviation. For SR and SH the values of the parameter are the same, within the standard deviation, and they are as high as 1.8 s at 500 Hz and 1.7 s at 1 kHz. Although these band values are comparable with the respective results in AS, the value at 2 kHz is now lower probably due to air absorption as discussed above. For this reason the "midaverage" (500 Hz, 1 kHz, and 2 kHz) is a little less than in AS and close 1.6 s. When SG is considered, the values of T_{20} are much lower and close to 1 s.

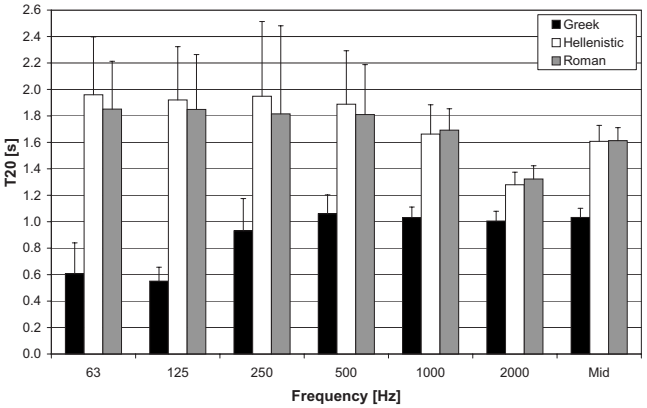


FIG. 9. The reverberation time in the three configurations of the Siracusa theater.

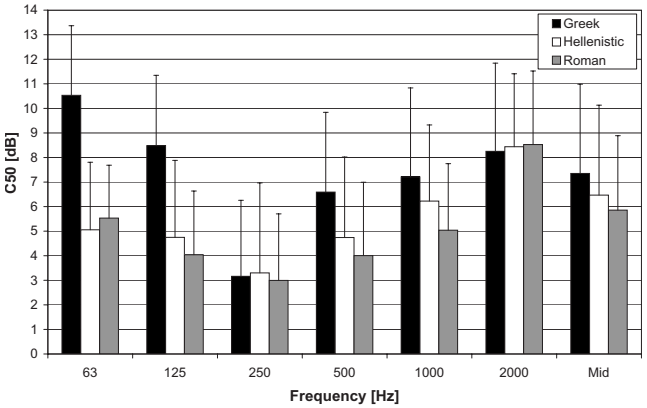


FIG. 11. Clarity for speech measured in the three configurations of the scale model.

The early decay time (EDT) parameter, not shown, presents lower mean values and with a very marked variability. In fact, the reflection coming in the very first part of the impulse response from the stage building arrives with different delays depending on the source position and this affects EDT rather than T_{20}. Thus in the present context T_{20} is more suited than EDT to achieve a robust qualification of the measured theaters according to the design typology. In other words the intrinsic EDT variability with position could mask the effect of architectural elements.

A further investigation was done exploiting the possibility of adding and removing the model parts. The results are reported in Fig. 10. The base model was SH (whose cavea has a constant slope) and the diazoma was suppressed with further tiers of steps. The base stage was also that of SH and four configurations of the model were fixed by combining the presence or the absence of the stage building and of the upper columned gallery.

From Fig. 10 it is evident that longer reverberation times are due to the presence of the upper gallery. In fact, the results show that removing the upper gallery causes the T_{20} at middle frequencies to decrease by about 25% whereas without the stage building the reverberation decreases by 15% at the same frequencies. In addition, while the removal of the stage does not affect the frequency trend of the reverberation time, when the gallery is removed one can find a more marked decrease in T_{20} in the lower frequency range. The upper gallery is thus mostly effective in this frequency range. Finally, the cavea alone, that is the theater without stage and gallery, has a T_{20} lower than the complete theater by 50%.

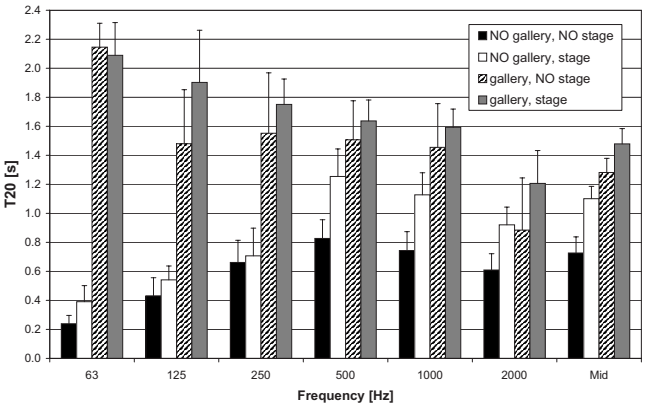


FIG. 10. Reverberation time measured in the scale model for four "hybrid" assembled configurations.

B. Clarity

The parameter clarity for speech C_{50} is reported in Fig. 11. It can be easily seen that, as for T_{20}, the values are very similar for SR and SH and are different for SG. Interestingly, the clarity in SR and SH is high although the T_{20} is also rather high. The frequency trend of C_{50} shows a dip in the 250 Hz octave band for all theaters which is actually more marked in SG.

To investigate what was producing the gap in the energy parameters, proper measurements were taken in the simplest configuration of the theater: SG without the stage building. The spark source was placed near the center of the orchestra and the microphone in a central sector in the middle of the cavea. In Fig. 12 the frequency analysis results of four different measures are plotted. The first (black dotted) is the

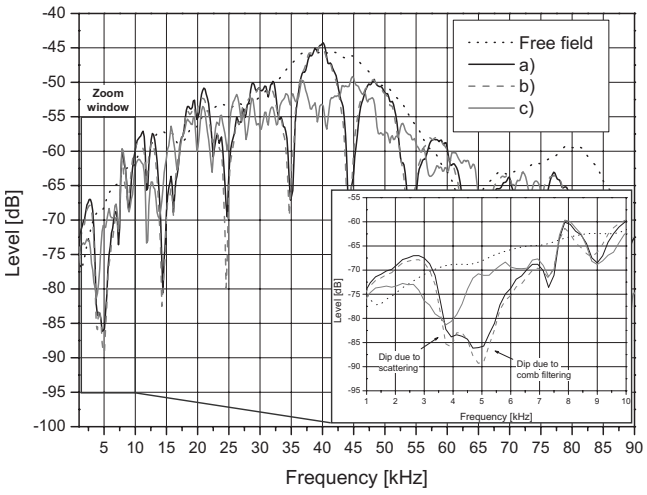


FIG. 12. Frequency analysis at the receiver position, in the model scale, showing the effect of comb filtering and a zoom of the curves from 1 to 10 kHz. (a) Greek theater configuration, (b) sound absorbing material close to the microphone, and (c) sound absorbing material covering the orchestra.

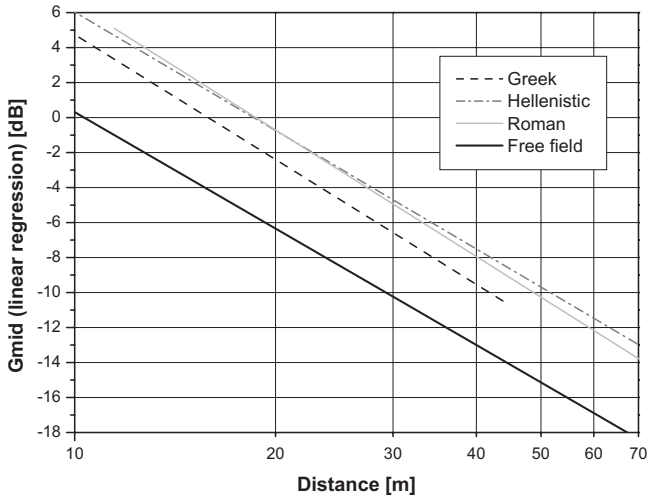


FIG. 13. Strength measured in the three configurations of the scale model compared with the free field.

frequency response of the spark source in a free field; this curve is reported only for comparison and its level is not related to the other curves. Curve (a) (black solid) is the reference configuration described above. It appears that the reason for the dip at 250 Hz (real scale) in the clarity graphs is due to the comb filtering generated by the interference between the direct sound and the reflection from the orchestra floor, which has a remarkable dip at 5 kHz (model scale). Then to exclude other effects generated very close to the microphone, the capsule was surrounded by sound absorbing material. Curve (b) (gray dashed) shows the result of this measurement: it can be seen that this curve is very similar to curve (a) representing the reference configuration. Only the frequency gaps caused by comb filtering are a bit more pronounced. Finally, as a check, the absorption around the microphone was removed and a layer of the same sound absorbing material was laid to cover the whole orchestra. The result is described by curve (c) (gray solid), which shows a course similar to the free field condition for most of the frequency range. Actually the dip at 5 kHz (250 Hz real scale) is removed but, very interestingly, curve (c) shows another dip in the frequency region close to 4 kHz (200 Hz real scale). This second dip is not removed neither by placing sound absorbing material on the orchestra nor very close to the microphone. A further inspection on the geometry of the cavea revealed that the half wavelength of this frequency corresponds exactly to the diagonal period of the steps.

Moreover the first dip at 5 kHz is presented here for a single position and moving the microphone in the cavea would cause just a minor frequency shift. For this reason this effect is also visible in the average value of the energy parameters in the 250 Hz band real scale. Finally, this comb filtering effect is less marked in SH and SR since more reflections coming from the bigger stage buildings and from the upper gallery reach the receiver and add in an uncorrelated way.

C. Sound strength

In Fig. 13 the strength G_{mid} propagation curves are reported. As above they are expressed with the regression lines for each data set to make the graph more intelligible and, for a better comparison, only the values measured with the source in the orchestra are shown. The receivers were placed on a line along a central sector of the theater. The propagation curve for the source in the free field is also reported for reference.

TABLE II. Linear fit parameters for measured theater sound level decays $[G_{\mathrm{mid}}=A_i+B_i \log_{10}(\mathrm{dist})\ (\mathrm{dB})]$.

Theater	A_i (dB)	B_i (dB)	R
Free field	22.41	−22.10	−0.99
Roman	30.54	−24.02	−0.99
Hellenistic	28.52	−22.49	−0.99
Greek	28.46	−23.72	−0.99

It is to be noted that, different from the case of *in situ* measurements, it is impossible to neglect the air absorption when calculating the free field propagation curve. Consequently, a MATLAB code was written to calculate the level of a source in a free field at a certain distance by introducing the correction for the air absorption given in the standard ISO9613-1.[19] This modified free field is taken as the reference propagation curve for the evaluations. The measurements relative to the G parameters in the three theaters were taken under the same temperature and humidity which were, respectively, 15 °C and 35%. Using these data the G parameter was calculated first for single frequencies in steps of 0.5 Hz, then for the 10 kHz, 20 kHz, and 40 kHz octave bands and finally averaged to obtain G_{mid}. The distances were then rescaled by 20 times. The linear fit $Y=A_i+B_iX$, where $Y=G_{\mathrm{mid}}$ and $X=\log(\mathrm{distance})$, was calculated for all the data sets. For the free field the results are $A_{\mathrm{free}}=22.41$ dB and $B_{\mathrm{free}}=-22.10$ dB with a regression coefficient $R=-0.99$. All the parameters calculated with the linear fits are summarized in Table II. In general the G_{mid} values are quite low compared to closed theaters and the trend constantly decreases with a slope very close to the free field.

With no surprise, since SR and SH are very similar, their sound strength propagation curves are quite close to each other and are about 5 dB higher than the free field values in the considered range of distances. Moreover, the 1 dB lower values found in SG are probably due to a less efficient reflection from the stage building caused by the presence of the portico.

VII. THE ACOUSTIC EVOLUTION OF ANCIENT THEATERS

The main characteristics of the theaters included in this work are summarized in Table III. The above set of scale model measurements can be compared with the data measured *in situ* reported in Sec. III. As is evident from the acoustical measurement results, derived from real theaters (Figs. 2, 3, and 5) and from the scale model configurations (Figs. 9, 11, and 13), the sound field in the ancient theaters has a unique character. The contributions of the specific architectural parts can be traced back in the impulse response (Fig. 8). In particular, the orchestra floor and the stage build-

TABLE III. Main characteristics of all the theaters investigated in this work.

Theater	Gallery	*Cavea* diameter (m)	*Cavea* slope	Stage building	Stage building height (m)	RT_{mid} (s)	Seating capacity
Siracusa Roman	Yes	143	20°–25°	Yes	20.0	1.63	15 000
Siracusa Hellenistic	Yes	143	20°	Yes	15.4	1.61	15 000
Aspendos	Yes	98	33°	Yes	26.0	1.68	15 000
Taormina	No	109	39°	Yes (part.)	20.0	1.16	5 500 (today)
Siracusa Greek	No	85	20°	Yes	8.6	1.03	5 000
Jerash	No	63	43°	Yes (part.)	8.5	1.19	3 000 (today)
Siracusa Greek–No stage	No	85	20°	No	⋯	0.40	5 000
Segesta	No	63	26°	No	⋯	0.45	4 000
Delphi	No	50	28°	No	⋯	0.50	5 000

ing provide very strong early reflections and the cavea contributes with the scattered sound that makes up the reverberant tail. Anyway, even if the theaters have no ceiling, which can be replaced by an ideal surface of unit absorption, the reverberation time in Roman theaters is similar to closed theaters, but in this case the clarity is higher and the sound strength is very low.

The nature of the scattering in the tail deserves further specific investigations. In fact, for a correct interpretation of the phenomenon it is necessary to resort to an advanced wave scattering theory as reported in Ref. 10 but the present measurements cannot be directly matched to what theoretically predicted in Ref. 10. A new dedicated set of measurements in the scale model is required to exactly correlate the wave theory of scattering with the experimental results. At present, scattering measurements on a sample of steps of the scale model have been accomplished using advanced techniques.[20] It was found that the typical "knee" of the resulting scattering curve for the step layout lies around 200 Hz. This frequency region coincides with the frequencies where, as evidenced by curve (c) of the zoom window in Fig. 12, the deep was originated by the cavea steps.

The effect of the architectural elements can be initially tested through examination of the measured RT_{mid}, the average of the reverberation time (T_{30} for the real theaters and T_{20} for the scale model) in the 500 Hz, 1 kHz, and 2 kHz octave frequency bands, also reported in Table III. The results of such procedure are shown in Fig. 14 where different grays correspond to the presence of the various architectural parts, the full bars refer to real theaters, and the slashed bars to the scale model configurations. Three groups can be identified: the first including SR, SH, and AS, the second with TA, SG, and JE, and the last with SG without stage, SE, and DE.

From this graph (Fig. 14) it can be seen that the RT_{mid} values can be grouped together according to their architectural layout. In fact, the measured RT_{mid} in the AS theater almost matches those measured in SR and SH, having both the complete stage building and the upper gallery. Furthermore, it can be noted that, despite the slight slope differences and the spread of overall dimensions, the measured RT_{mid} is similar. Finally, as already noted, the RT_{mid} in SH and SR is similar even if they have very different stage buildings.

Moving to the second group of data, the theaters without the upper columned gallery, the JE and TA theaters have a RT_{mid} very similar to those measured in SG, even if they had different dimensions of the cavea and of the stage building. Theaters in the last group show almost identical RT_{mid}.

Similar considerations come out from the analysis of clarity parameter averaged for the middle frequencies, $C50_{mid}$. In Fig. 15 the average value measured for the three groups is shown. Even though the clarity has a much higher variability with respect to the receiver positions, the three groups of theaters can be still identified. Analyzing in detail Fig. 15 it can be seen that the differences between the theaters with and without the gallery are very small while, when removing the skene wall, $C50_{mid}$ becomes very high.

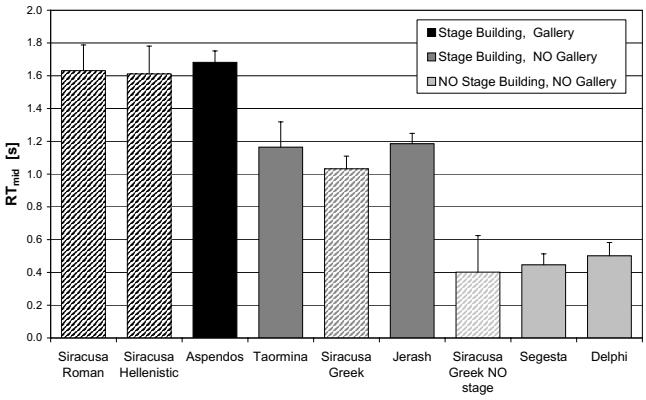


FIG. 14. Comparison between the RT_{mid} measured in all the analyzed theaters. Full bars: real theaters, slashed bars: scale model.

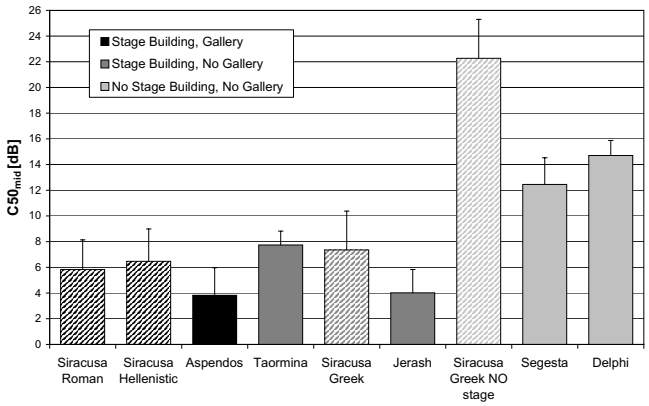


FIG. 15. Comparison between the $C50_{mid}$ measured in all the analyzed theaters. Full bars: real theaters, slashed bars: scale model.

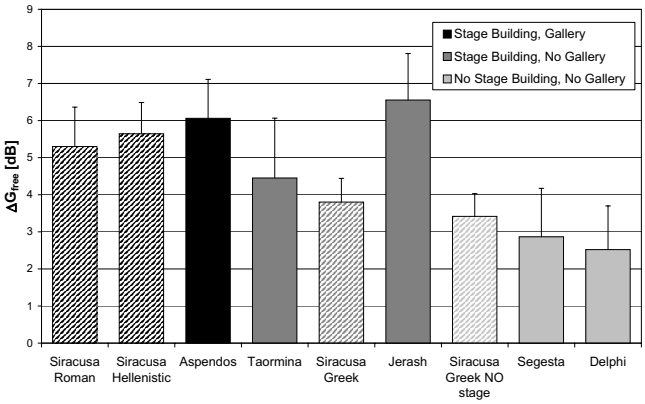


FIG. 16. Comparison between the ΔG_{free} (average difference between G_{mid} in the theaters and in a free field) measured in all the analyzed theaters. Full bars: real theaters, slashed bars: scale model.

Finally, the averaged difference ΔG_{free} between the linear fit curve of the G_{mid} and the free field can be analyzed.

The differences calculated for all the theaters are reported in Fig. 16. The bars represent the standard deviation from the mean value and in this case they take into account also the slope difference between the sound propagation curves in the theater and in the free field.

The strength depends on the stage wall providing a first reflection and on the upper gallery providing a more delayed sound but also on the slope of the cavea as the case of the steep JE clearly indicates. The interplay and ranking of these factors are not straightforward. For instance, one can argue that the stage wall reflection is surely more valuable for most listeners than a delayed reflection coming from the gallery behind. In this respect, it can be safely stated that the increased stage building dimensions in Roman theaters helped the listeners with the sound strength more than the upper galleries. Moreover if JE is compared to the first group (SR, SH, and AS) it is found that a steep slope is able to increase the strength at the listeners at least as does a complete stage wall. This is consistent with the cited strategy used by ancient designers to optimize the listening conditions at farthest seats by increasing the final slope. Regarding the comparison of the first and the second group it is also to be noted that, should JE be taken out of its group, a clear decreasing trend could be reported from group one to the next. This means that in many cases, at least in theaters where the slope does not exceed 25°–30°, the contribution of the upper gallery could be precisely isolated. Finally the complete removal of the stage wall reflection which characterizes the third group (SG without stage, SE, and DE) causes the G to decrease significantly. These theaters can be regarded as a sort of basic acoustical spaces, since the sound strength profits of just one single floor reflection, and the ΔG_{free} is consistently assessed at values very close to 3 dB. It is to be stressed here that a similar result could never be obtained without an appropriate sloping of the audience because of the screening effect of the rows in front.

VIII. CONCLUSIONS

The above comparisons are a confirmation that the evolution of acoustics in ancient theaters can be qualified by architectural details and can be quantified by acoustical measurements. In this respect a few elements were isolated, namely, the stage wall, the upper columned gallery, and the slope of the cavea. Their interplay seems capable of describing the overall acoustical properties of several types of ancient theaters even if their design does not strictly match the model configurations. For this reason a summary of architectural features and the related RT_{mid} for measured theaters is reported in Table III. Moreover the findings allow to propose a set of expected values for the main acoustical parameters of theaters falling in the three groups identified above based on their design elements. The proposed values are summarized in Table IV.

The reverberation time has a limited variability so the estimate of this parameter in other theaters with similar design should be rather accurate. In this respect it was also possible to rank the different elements considering how they affect the reverberation in an ancient theater. In fact, it was found that the overall space dimensions seem to have less importance than the completeness of the theater itself in its different parts. This behavior differs from the classical theory of reverberation in closed spaces, where the reverberation is directly proportional to the room volume. Moreover, the upper gallery seems to be the chief element to keep the sound energy circulation inside the cavea prolonging the reverberant tail, whereas the slope of the cavea is not important for the reverberation time. Unfortunately, the role of occupancy still remains unexplored and shall be possibly considered in the future.

As regards ΔG_{free} it can be seen that when passing from group one to group three the strength is halved since one of the two reflections (that on the *skene* wall) is lost. Anyway, as discussed above, the parameter is more affected by the state of conservation of the stage building, and the steepness of the *cavea* plays an important role too. For this reason the intermediate type of theaters (second row of Table IV) carries most of the variability for this parameter.

TABLE IV. Values of the main acoustical parameters expected in ancient theaters.

Type	Real	Model	RT_{mid} (s)	$C50_{\text{mid}}$ (dB)	ΔG_{free} (dB)
Complete Roman	Aspendos	Siracusa Roman and Hellenistic	1.64 ± 0.13	5.4 ± 2.3	5.7 ± 1.0
Roman or Greek without gallery and incomplete stage building	Taormina Jerash	Siracusa Greek	1.13 ± 0.10	6.4 ± 2.0	4.9 ± 1.2
Greek without stage building	Delphi Segesta	Siracusa Greek without stage building	0.45 ± 0.12	16.5 ± 2.1	2.9 ± 1.0

Moreover, the values of clarity are mostly governed by the reflection of the skene. The parameter slightly changes when removing the gallery but, when the stage wall is removed, its values more than double.

The acoustical measurements also showed that in ancient theaters the wave effects cannot be neglected. In particular, comb filtering effects have been discussed due to the orchestra floor reflection but more subtle wave phenomena are caused by the scattering of the cavea. For their full understanding further experimental and theoretical researches are required.

Finally, as discussed above, the work has adopted the standard in Ref. 14, whose objective acoustical parameters with their related optimal ranges were developed for closed spaces. Here some unusual behavior was outlined when employing the same concepts in open air theaters. Thus the subjective significance of the present findings and the eventual adoption of peculiar objective parameters are still a matter of research which will be addressed in the near future.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. T. Hidaka of Takenaka Corporation, Chiba, Japan for kindly lending the miniaturized dodecahedron source. Dr. Shin-ichi Sato of Hangyang University, Korea and Dr. Hiroyuki Sakai are also acknowledged for support during acoustical measurements in Taormina and Delphi. The measurements in Segesta were taken within the Italian National Research Project "PRIN anno 2005-prot. 2005081870."

[1]G. Izenour, *Theater Design*, 2nd ed. (Yale University Press, New Haven, CT, 1996).
[2]F. Canac, *L'Acoustique des Theatres Antiques: Ses Enseignements (The Acoustics of Ancient Theatres: Their Teaching)* (CNRS, Paris, 1967).
[3]L. Cremer and H. Muller, *Principles and Applications of Room Acoustics* (Applied Science, London, 1982).
[4]B. Papathanasopoulos, "Acoustiche messungen in theater von Epidauros (Acoustical measurements in the theatre of Epidauro)," Proceedings of Fifth ICA, Liege, 1965.
[5]R. Shankland, "Acoustics of Greek theatres," Phys. Today **26**(10), 30–35 (1973).
[6]D. Irkli, "Acoustical properties of ancient theatres: Computer simulation and measurements," Ph.D. thesis, Middle East Technical University, Ankara, Turkey (1995).
[7]S. Sato, H. Sakai, and N. Prodi, "Acoustical measurements in ancient Greek and Roman theatres," Proceedings of Forum Acusticum, Sevilla, 2002.
[8]S. Vassilantonopoulos and J. Mourjopoulos, "A study of ancient Greek and Roman theater acoustics," Acta. Acust. Acust. **89**, 123–136 (2003).
[9]D. Noson, T. Bosworth, and N. Folk, "Private communication on the 1:10 scale model of delphi theatre at the college of architecture and urban planning, University of Washington" (2000), URL: http://www.caup.wahsington.edu/support/ newsletter/2000_winter.pdf (last viewed 7/11/08).
[10]N. Declercq and C. Dekeyser, "Acoustic diffraction effects at the hellenistic amphitheater of epidaurus: Seat rows responsible for the marvelous acoustics," J. Acoust. Soc. Am. **121**, 2011–2022 (2007).
[11]P. Chourmouziadou and J. Kang, "Acoustic evolution of ancient Greek and Roman theatres," Appl. Acoust. **69**, 514–529 (2007).
[12]L. Polacco, *Il Teatro Antico di Siracusa (The Ancient Theatre of Siracusa)* (Maggioli, Rimini, 1981).
[13]"Erato project (identification evaluation and revival of the acoustical heritage of ancient theatres and odea), ica3-ct-2002-10031. Part of the European Commission Fifth Framework Incomed Programme," URL: http://server.oersted.dtu.dk/www/oldat/erato/ (last viewed 7/11/08).
[14]"ISO/dis 3382-1:2004. Acoustics—Measurement of room acoustic parameters—Part 1: Performance rooms," Technical Report, ISO, 2004.
[15]H. Kuttruff, *Room Acoustics*, 4th ed. (E&FN SPON, London, 2000).
[16]N. Xiang and J. Blauert, "Binaural Scale Modelling for Auralization and Prediction of Acoustics in Auditoria," Appl. Acoust. **38**, 267–290 (1993).
[17]K. Lorenz, "Auralisation of a scale model of the Royal Albert Hall," J. Acoust. Soc. of the Netherlands **158** (**2001**).
[18]A. Farnetani, "Investigation on the acoustics of ancient theatres by means of modern technologies," Ph.D. thesis, Università degli Studi di Ferrara, Ferrara, Italy (2006).
[19]"ISO 9613-1:1993. Acoustics—Attenuation of sound during propagation outdoors—Part 1: Calculation of the absorption of sound by the atmosphere," Technical Report, ISO, 1993.
[20]A. Farnetani, N. Prodi, and R. Pompoli, "Measurements of sound scattering of the steps of the cavea in ancient open air theatres," Proceedings of the International Symposium on Room Acoustics, Sevilla, 2007.

Experimental and numerical investigations on melamine wedges

S. Schneider[a)]
CNRS/LMA, 31 chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France

(Received 18 January 2007; revised 19 March 2008; accepted 23 May 2008)

Melamine wedges are often used as acoustic lining material for anechoic chambers. It was proposed here to study the effects of the mounting conditions on the acoustic properties of the melamine wedges used in the large anechoic chamber at the LMA. The results of the impedance tube measurements carried out show that the mounting conditions must be taken into account when assessing the quality of an acoustic lining. As it can be difficult to simulate these mounting conditions in impedance tube experiments, a numerical method was developed, which can be used to complete the experiments or for parametric studies. By combining the finite and the boundary element method, it is possible to investigate acoustic linings with almost no restrictions as to the geometry, material behavior, or mounting conditions. The numerical method presented here was used to study the acoustic properties of the acoustic lining installed in the anechoic chamber at the LMA. Further experiments showed that the behavior of the melamine foam is anisotropic. Numerical simulations showed that this anisotropy can be used to advantage when designing an acoustic lining. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2945166]

PACS number(s): 43.55.Pe, 43.55.Ev, 43.58.Bh, 43.55.Wk [KAC] Pages: 1568–1576

I. INTRODUCTION

To perform acoustic measurements, it is of crucial importance to have information about the characteristics of the environment in which the measurements are performed. As it is not always feasible to carry out experiments in a real free space, most of these measurements are usually performed in an artificial free space consisting of either an anechoic or a semianechoic chamber. The quality of these chambers, i.e., the sound levels reflected by the walls, has been studied since the early 1940s.[1–6] The most decisive factor on which the quality of an anechoic chamber depends is the acoustic lining of the chamber. Much research has therefore focused on optimizing the acoustic lining by finding the optimum geometry, size, and material of the lining. This research has involved experimental studies on various linings[7–9] and a comprehensive parametric study using numerical simulations. Most of these numerical simulations have involved either analytical models such as the double porosity model[10,11] or numerical methods such as the finite element method (FEM),[12–14] the boundary element method (BEM),[15] and the boundary-condition-transfer algorithm.[16] In most of these studies, a fairly simplified geometry and/or material model has been used. In particular, the frame of the absorbing material has often been assumed to be motionless. In this study, we present a numerical approach that imposes very few restrictions on the geometric shape and the arrangement of the components of the lining, their mounting conditions, and the material model used. This was achieved by coupling the FE and BE methods. Since the FEM allows great flexibility as far as the geometry, boundary conditions, and material models are concerned, it was used to model the absorbing material and its mounting conditions. The fact that the FEM requires the entire volume to be discretized is a disadvantage with this method, especially when one has to mesh two volumes with a complex geometry sin such a way that the two meshes match at their common interface. To overcome this difficulty, the BEM was used to model the sound propagation in the surrounding homogeneous fluid. The surface mesh required by this numerical method was generated on the fluid/lining interface based on the given FE mesh. The efficiency of the BEM itself was increased by using a multilevel fast multipole algorithm (MLFMA). The resulting numerical method was used to investigate the acoustic behavior of the acoustic lining of the large anechoic chamber at the Laboratoire de Mécanique et d'Acoustique (LMA) in Marseille, the inner dimensions of which are $5.4\times6.3\times11.4$ m^3, measured from wedge tip to wedge tip. The present study was motivated by the imperfections observed in the anechoic chamber above its cutoff frequency. Here, we focus on the contribution of the mounting to the acoustical performance of the lining. Studying the behavior of a complex geometry exposed to plane wave excitation is the first step toward the numerical simulation of an entire anechoic chamber in which the real geometry of the acoustic lining is taken into account.

The remainder of this paper is organized as follows. Section II gives the results of impedance tube measurements performed on the melamine wedges, neglecting the iron wire frame. The influence of this frame on the absorption coefficient is studied experimentally in Sec. III. Section IV deals with the numerical simulation of the experiments and with the influence of the material model used to describe the behavior of the sound absorbing material, and discusses possible means of improving the acoustic lining.

II. IMPEDANCE TUBE MEASUREMENTS I

To determine the absorption coefficient of the acoustic lining used in the large anechoic chamber at the LMA, impedance tube measurements were carried out. The acoustic

[a)]Electronic mail: schneider@lma.cnrs-mrs.fr.

lining consists of 3720 melamine wedges with a tapering section 0.7 m in length. The base of the wedges is 0.4 m long and has a rectangular cross section measuring 0.3 $\times 0.3$ m^2. The individual wedges are held in place by an iron wire frame. The large impedance tube at the LMA measuring $1.2 \times 1.2 \times 6$ m^3 was used to examine the acoustic properties of samples consisting of 4×4 wedges without the iron wire frame. In situations of this kind, accurate positioning of the sample is required. The wedges must be placed properly at the end of the tube without leaving any air gaps between the wedges or between the walls and the wedges, as air gaps affect the results of the measurements.[17] To meet this requirement, the sample was first clamped outside the tube using four thin wooden plates. Once the clamped sample had been placed in the tube, the clamping was released, leaving the wooden plates in the tube. This setup also made it possible to place the sample at various distances from the end of the tube in order to carry out measurements with and without an air gap behind it.

The reflection coefficient of this sample was measured in the 20–200 Hz frequency range using the transfer-function method.[18–20] For this purpose, a 1 in. microphone was used to record the sound pressure at the center of the cross section of the tube at various distances from the melamine wedges. A second microphone was used as the reference microphone to check the stability of the experiment. Both microphones were connected to a Nexus preamplifier connected to a Stanford SR 785 signal analyzer. To increase the signal-to-noise ratio, a pure sine tone was used as the acoustical excitation. The signal analyzer was run in the synchronous-demodulation mode to eliminate the influence of frequencies other than the excitation frequency. Data were postprocessed using the SCILAB 2.7 software program. The sound pressure was recorded at distances $z_i \in [0.7, 0.85, 1.0, 1.15, 1.3, 1.45]$ m from the sample. This range of distances was chosen to ensure that plane waves were established at all the microphone positions. At each pair of positions (z_n, z_m) with $\Delta z = z_m - z_n = 0.3$ m, the reflection coefficient R was calculated as follows:

$$R_{nm} = \frac{e^{jkz_n} - h_{nm}e^{jkz_m}}{h_{nm}e^{-jkz_m} - e^{-jkz_n}}, \tag{1}$$

with the transfer function $h_{nm} = p_n / p_m$ of the corresponding sound pressures. Almost no differences were observed between the reflection coefficients R_{nm} calculated in all the pairs investigated. However, the reflection coefficient R of the sample was determined using the sound pressures recorded at distances $z_n = 1$ m and $z_m = 1.3$ m from the sample. Based on the reflexion coefficient R, the absorption coefficient $\alpha = 1 - |R|^2$ and the specific surface admittance $Y = (1 - R)/(1 + R)$ were calculated. The absorption coefficients α of the sample of four wedges measured with and without an air gap behind the sample are given in Fig. 1. The known effect of an air gap behind an absorbing layer, which results in higher absorption at low frequencies and less absorption at higher frequencies,[21] can be clearly seen here. Based on the results of this experiment, the cutoff frequency of an anechoic chamber equipped with these melamine wedges was estimated to be around 80 Hz, since at this frequency,

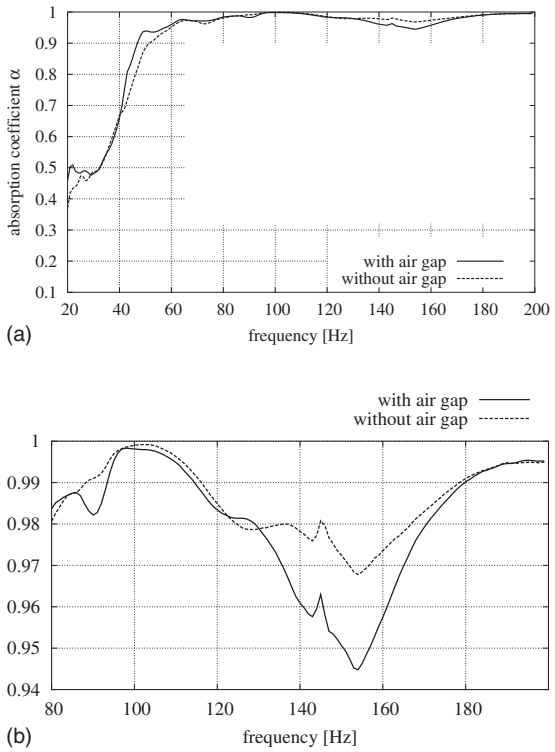


FIG. 1. Absorption coefficient of a 4×4 sample of melamine wedges measured with and without an air gap behind the sample (sub-figure (a)). Wedges were placed in the impedance tube without the iron wire frame. The subfigure (b) shows a zoom in the 80–200 Hz frequency range.

the absorption of the lining is greater than 99%. However, no attention has been paid so far to the fact that in the 120–180 Hz frequency range, the absorption coefficient decreases to below 98%, reaching a minimum of 95% at 155 Hz, although perturbations in the chamber have been reported[22] to occur in this frequency range. Recent measurements performed in the 20–200 Hz frequency range in the anechoic chamber also showed that above the cutoff frequency, deviations from the free-field conditions larger than the allowed 1.5 dB (ISO 3745 and ANSI S12.35) occur. To determine these deviations, the sound pressure radiated by a loudspeaker was recorded along a line parallel to a wall with a spatial resolution of 0.2 m. From the sound pressure p_{exp} measured at each microphone position and the pressure p_{inc} radiated by the sound source, the deviation from free-field conditions L_p given by $L_p = 20 \log(|p_{\mathrm{exp}}|/|p_{\mathrm{inc}}|)$ was calculated. The incident sound pressure p_{inc} generated by the sound source at the corresponding microphone position was determined using a numerical model for the source.[23] Figure 2 shows the deviations obtained with two selected microphone positions, 1.6 and 3.1 m from the sound source. The perturbations occurring at 120 and 170 Hz can be only partly explained by the absorption coefficient of 98% measured at these frequencies. One of the reasons for this failure may be that the mounting conditions of the wedges applied in the experiment differed from those used in the anechoic chamber. A second impedance tube experiment was therefore carried out in which the influence of the iron frame holding the wedges in place when they are installed in the anechoic chamber was investigated more closely.

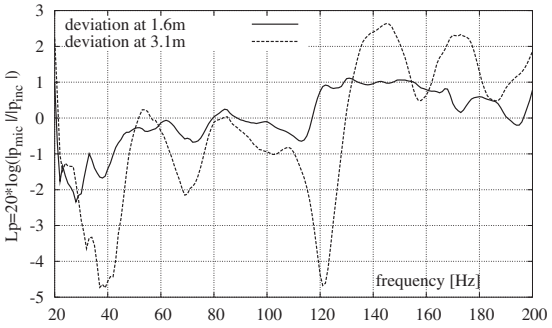


FIG. 2. Deviation from free-field conditions observed in the anechoic chamber at the LMA.

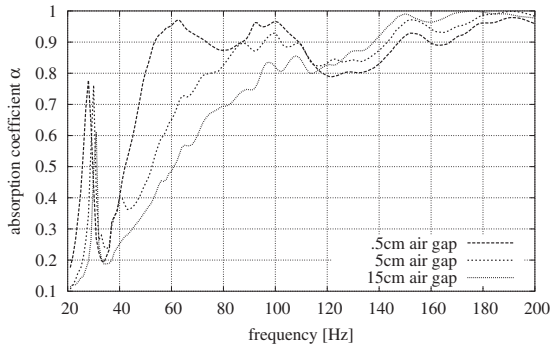


FIG. 3. Absorption coefficients measured with various air gap sizes around the sample of absorbing material.

III. IMPEDANCE TUBE MEASUREMENTS II

Several previous studies[17,24–26] have shed light on the influence of the boundary conditions on the acoustic behavior of absorbing materials. In the present context, this involves the boundary conditions themselves when experiments are carried out to determine the acoustic properties of a material, as well as the mounting conditions of an absorbing material in real life applications. Studies on the acoustic properties of a lining based on experiments and/or numerical simulations should therefore not be performed under mounting conditions that differ from those present in the applications. The usefulness of the absorption coefficient presented in the previous section for the characterization of an anechoic chamber is therefore questionable. A second impedance tube measurement was therefore carried out. The point of this experiment was to study the influence of the iron wire frame on the absorption coefficient for normal incidence of plane waves and to obtain the experimental data subsequently used to confirm the validity of the numerical models used in Sec. IV. Note that the absorption coefficient measured cannot be used to assess the lining of the anechoic chamber, as the experimental configurations used differed from that of the chamber. However, experimental data can be used to study the role of the mounting condition on the acoustical properties of the wedges.

To account for the mounting conditions of the melamine wedges in the anechoic chamber at the LMA, the same wire frame was used in a second impedance tube experiment to hold the wedges in place. An iron wire frame holding 3×3 wedges was used and the frame was screwed to three wooden bars, leaving an air gap of 11 cm behind the sample. This air gap was slightly smaller than that existing in the anechoic chamber. The wooden bars were assumed to be motionless, which was confirmed by measuring their displacement. A sample consisting of 3×3 wedges leaves an air gap of 15 cm around the melamine wedges, thus preventing any lateral contact between the absorbing material, the iron wire frame, and the walls of the impedance tube. This configuration is characterized by well defined boundary conditions and can therefore be used to study the influence of the wire frame. In addition, these clearly defined boundary conditions rule out a possible source of error in the subsequent numerical analysis. In order to obtain experimental data on various configurations, the air gap around the sample was closed using a multicomponent wooden frame (see left subfigure in Fig. 6). Three different configurations were studied:

(1) a wooden frame measuring 14.5 cm in width, leaving an air gap of approximately 0.5 cm between the sample and the wooden frame;
(2) a wooden frame measuring 10 cm in width, leaving an air gap of 5 cm between the sample and the wooden frame;
(3) no wooden frame, leaving an air gap of 15 cm between the sample and the walls of the tube.

A fourth possible configuration in which the air gap remaining in configuration one was closed up with an elastic seal was not studied, as the presence of the sealing introduces additional uncertainties about the boundary condition at the point of contact between the sealing and the wedges.

The first configuration was installed and measurements were performed. The wooden frame was then removed to obtain configurations 2 and 3 without changing the position of the sample of absorbing material. Again, the transfer-function method and the same experimental setup as in Sec. II were used to measure the acoustic properties of the melamine wedges. The absorption coefficients measured in the three configurations are given in Fig. 3. Decreasing the size of the air gap led to a significant increase in the absorption coefficient at low frequencies and to a slight decrease in the absorption coefficient at frequencies above 120 Hz. Besides this effect, an increase in the absorption was observed in the 80–110 Hz frequency range when the size of the air gap was reduced. This increase was attributed to longitudinal structural modes of the wedges. These modes were identified by measuring the longitudinal displacement of the wedge tips using an MTI-2000 fiber-optic displacement sensor. The displacements measured normalized by the pressure at the corresponding measurement points are given in Fig. 4. This figure also indicates the numbering of the wedges used in the following discussions. A mode can be seen to occur at 30 Hz, where the entire sample vibrated in the wire frame. This mode caused a sharp peak in the absorption coefficient at the corresponding frequency (see Fig. 3). A mode involving wedges 2 and 5 at around 65 Hz caused a slight increase in the absorption coefficient at all three air gap sizes in Fig. 3. At 85 Hz, we observed a mode involving wedge 5, at 95 Hz a mode involving wedges 1 and 2, and at 102 Hz a mode

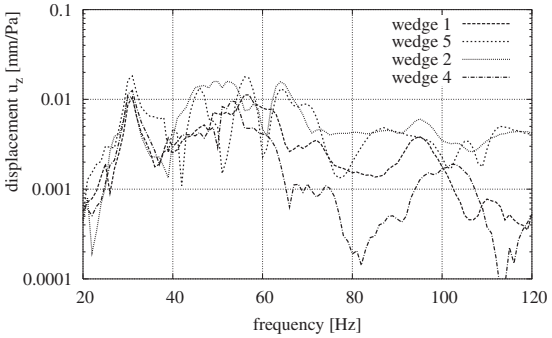


FIG. 4. Longitudinal displacement of the melamine wedges measured with a fiber-optic sensor (left subfigure). Numbering of the wedges (right subfigure).

involving wedge 6. These modes were responsible for the higher absorption coefficient observed in the corresponding frequency range of 80–110 Hz.

The absorption peak at around 60 Hz followed by a decrease in the absorption was very similar to the effect of double porosity observed by Sgard *et al.*[31] The latter authors also observed the influence of modes of the elastic frame of the absorbing material, which gave rise to further peaks in the absorption coefficient curve. It is worth noting the significant decrease in the absorption which occurred in the 110–140 Hz frequency range. In this frequency range, as well as in the 70–90 Hz range, there exist no structural modes that might lead to any additional absorption to that exerted by the material itself. Spreading out the structural modes of the wedges over the 70–150 Hz frequency range therefore provides a means of improving the quality of the acoustic lining. This will be discussed in Sec. IV.

Given the tendency for the absorption coefficients to decrease in the 110–180 Hz frequency range depending on the air gap size, it can be assumed that the breakdown of the absorption coefficient occurring in this specific frequency band will also occur when there exists no air gap. However, this configuration could not be investigated experimentally at the LMA, and numerical simulations needed to be performed to assess the quality of the acoustic lining used in the anechoic chamber in question.

IV. NUMERICAL SIMULATION

As explained in the previous section, the absorption coefficient of a sample of melamine wedges held in place by an iron wire frame without any air gap around the sample could not be determined in the impedance tube at the LMA, since the frame and the wedges constitute a sample that is slightly larger than the cross section of the impedance tube. Since the lateral contact between the sample and the wall of the tube gives a zero displacement condition at this position, the mounting conditions present in the anechoic chamber are not accurately simulated by this experimental configuration. The absorption coefficient for normal incidence of plane waves of a 3×3 sample including the iron wire frame was therefore determined using numerical methods. These simulations were performed in several steps. First, measurements were carried out in a smaller impedance tube on a sample with a simple geometry, and simulations were then performed to identify the parameters of the absorbing material. Second, the three experimental configurations presented in the previous section were simulated to check the validity of the numerical model and the material parameters with the complex lining geometry. Third, the absorption coefficient of the acoustic lining for the configuration used in the anechoic chamber at the LMA was determined numerically.

The numerical simulations were performed by coupling the FEM and the BEM. The absorbing material, the air gap around and behind it, and the iron wire frame were modeled using the FEM, whereas the sound propagation in the remaining part of the rectangular impedance tube was simulated using the BEM.

Sound propagation in the melamine foam was modeled using the Biot theory.[27,28] As the absorbing material will be coupled to a fluid and the pressure field therein will be of interest, the Biot theory formulated in the variables $\{u,p\}$ as described by Atalla *et al.*[29] was applied to model the displacement field u of the frame and the sound pressure p in the fluid phase of the porous material. For a more detailed description of the Biot model and the notation used, see Refs. 27–29 Sound propagation in the fluid domain Ω_F, i.e., the air surrounding the lower part of the wedges and the air gap behind the wedges, and the dynamic behavior of the iron wire frame were modeled using standard FE techniques. Hierarchic polynomials of order two were used as basis functions for the FEM. To describe sound propagation in the fluid domain Ω_{FT}, i.e., the air column in the impedance tube, the boundary integral representation of the solution of the time harmonic Helmholtz equation was used. The integral equation was solved using a collocation method with piecewise linear polynomials as basis functions. Because of the nonlocal support of the fundamental solution of the Helmholtz equation, the discretization of the integral equation yields dense matrices, giving the boundary element method a complexity of at least $\mathcal{O}(N^2)$ in a problem with N unknowns. This difficulty was overcome by applying a MLFMA[30,31] with which it is no longer necessary to explicitly calculate the dense matrices. Applying the MLFMA reduces the complexity of the BEM to $\mathcal{O}(N\log^2 N)$ Refs. 32–34 and makes this numerical method efficiently applicable to large scale problems. The subdomains were linked up via continuity conditions on the sound pressure and the normal surface velocity. The presence of the wooden frame was simulated by taking the fluid velocity to be zero at the corresponding positions.

Using the BEM to solve the sound propagation problem in the air column of the impedance tube abolishes the need to mesh this three-dimensional domain. The remaining three-dimensional domain Ω_B occupied by the porous material can be meshed much more efficiently when this can be done independently of the surrounding fluid domain. In addition, the geometry of the wedges can almost always be meshed using hexahedral elements rather than tetrahedral ones, which is preferable from the numerical point of view. Combining the FEM and the BEM makes it possible to deal with practically every possible wedge geometry and arrangement of the individual elements. The surface mesh required for the BEM, that is, the mesh of the surface of the much more

TABLE I. Material parameters of melamine as published in Ref. 36.

Material	Young's module E (N/m^2)	Poisson ratio ν	Density ϱ (kg/m^3)	Flow resistivity σ ($N\,s/m^4$)	Porosity Φ	Tortuosity α_∞	Characteristic dimensions Λ (m)	Characteristic dimensions Λ' (m)
melamine	0.16*E*+6	0.44	8.35	1.2*E*4	0.99	1.0	0.8*E*−4	0.28*E*−3

complex-shaped fluid domain, is generated directly by the nodes of the FE mesh on the contact surface, and hence no additional meshing is required for this part of the boundary Γ_{BE}. To simulate the impedance tube experiment, the walls of the tube obviously have to be meshed, however. The advantages of the FE/BE method used here will show up when simulating an entire anechoic chamber and modeling the real geometry of the acoustic lining,[35] as well as any radiating or scattering objects possibly present in the chamber. In this situation, all six walls of a rectangular chamber will have the geometry depicted in Fig. 6 with a total of 3720 wedges and an iron wire frame of a corresponding size. The fact that no three-dimensional fluid mesh is required to model the interior of the chamber will greatly simplify the numerical simulation.

Before the experiments presented in the previous section on melamine wedges with a complex geometry, the iron wire frame and the surrounding air were simulated. These simulations were performed on samples with a much simpler geometry to test the validity of the material parameters given in Table I. A loss factor $\eta=0.01$ was used to account for the structural damping of the frame of the melamine foam. This preliminary study was motivated by the fact that melamine foam has been reported to be anisotropic.[36] This anisotropy has been neglected so far and only the isotropic material parameters have been determined. Unfortunately, no samples of material were available that could be investigated in the large impedance tube in the 20–200 Hz frequency range, and a much smaller sample was therefore investigated in a smaller impedance tube but in a higher frequency range. To eliminate most of the errors liable to result from approximating the geometry of the sample, a cube of melamine foam measuring $6\times6\times6$ cm^3 was analyzed in a circular impedance tube with a diameter of 10 cm. Samples cut out of a wedge with the z-axis corresponding to the length of the wedge were glued onto the sample holder of the impedance tube. This configuration has well defined boundary conditions. The boundaries are free when the sample is in contact with air and motionless when it is in contact with the impedance tube. The absorption coefficients measured in the 300–1500 Hz frequency range with three sample orientations using the transfer-function method are given in Fig. 5 (upper left subfigure). It can be seen here that the peak in the absorption coefficient curve, which resulted from a resonance of the elastic frame, occurred at different frequencies with each of the directions. This shows that the material is not isotropic, which is in agreement with what has been reported in the literature.[24,36,37] The anisotropy may be attributable to the manufacturing process undergone by the material. The variations observed in the absorption curve, which were caused by resonances of the frame of the porous material, made it possible to determine the mechanical properties of the material in each of the three directions. Assuming that the material axes coincide with the cube's orientation means that an orthotropic material model can be used to describe the observed anisotropy. Based on the experimental data, at least the three Young's moduli can be determined, using a numerical model for the experiment. Here, only the FEM was used to simulate the experiment. Figure 5 gives the experimental data and the numerical results obtained using an orthotropic material model. The material parameters identified are summarized in Table II. The acoustic properties of melamine are assumed to be isotropic. As the experimental results were almost insensitive to the shear moduli, they were defined as $G_{ij}=E_i/2/(1+\nu_{ij})$. Experimental configurations that are sensitive to the shear modulus and Poison's ratio are being developed as part of an ongoing research project. The mechanical material parameters given in Table II and the acoustical parameters given in Table I, which resulted in good agreement between the experimental and simulated data in the case of this simple geometry, were used in the subsequent analysis of the melamine wedges.

Once the mechanical properties of the melamine foam had been identified, the experiments described in Sec. III were simulated by coupling the FE and the BE method, as described above. Numerical models for these experiments were therefore set up. Parts of the meshed geometry—wedges, air gap, iron wire frame and part of the BE-surface mesh—are shown in Fig. 6. The acoustical sound source, a loudspeaker with a diameter of 40 cm, was modeled by a uniform normal surface velocity at the corresponding end of the tube. This numerical model was used to determine the sound pressure at the same positions in the tube as in Sec. II, and the reflection coefficient was calculated using Eq. (1). As mentioned above, the material properties of the melamine foam are not isotropic, and attention must therefore be paid to the orientation of the material axis with respect to the geometry. decibel france, who installed the acoustic lining of the anechoic chamber at the LMA, stated that the wedges were cut out of the raw melamine blocks in the direction in which they were produced, or perpendicular to this direction in a random manner. However, the actual orientation of the material axes of the wedges used for the experiments could not be determined. To simulate this specific configuration, the sample of 3×3 wedges was modeled by setting Young's modulus in the direction of the wedge length to the values given in Table II for three wedges at a time. The configuration adopted here is shown in the left subfigure in Fig. 6, where each of the three material orientations is given by a

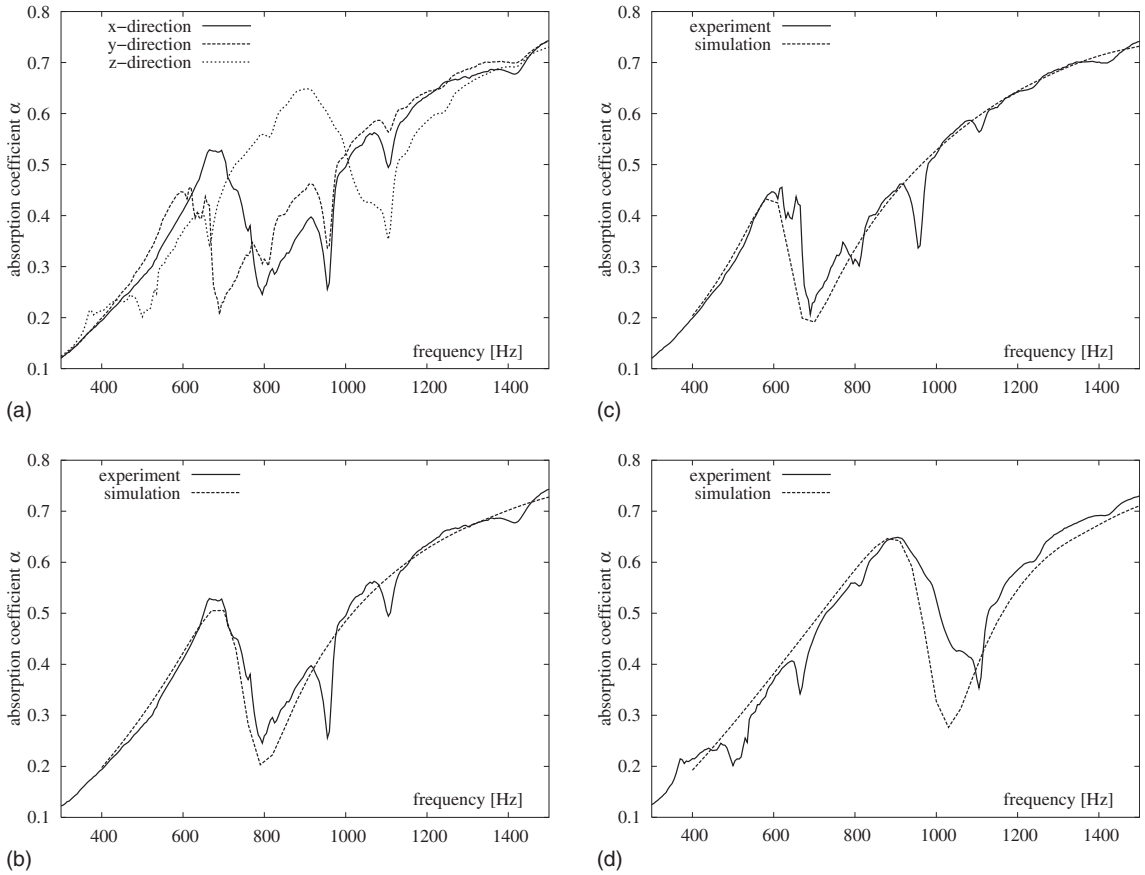


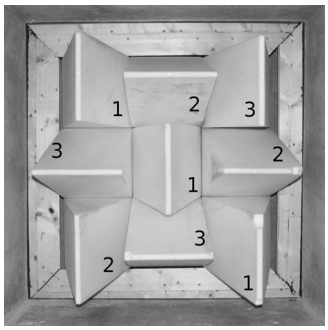
FIG. 5. Absorption coefficients of a $6\times6\times6$ cm^3 cube of melamine measured in different orientations of the cube (upper left sub-figure). Results of the numerical simulation of the experiment after fitting the material parameters to an orthotropic material model (upper right, lower left, and right subfigure).

different number. Numerical simulations were run on an SGI Origin3800 computer at the Center for Information Services and High Performance Computing at the Technische Universität Dresden, Germany.

The results of the numerical simulations can be compared with the experimental data obtained in the three different experimental configurations in Fig. 7. Simulation and experimental data agreed quite well over the entire frequency range when the Biot theory was used to model the absorbing material, although the exact position of the peaks could not be predicted in the case of a 15 mm air gap configuration. However, the numerical model gives a fairly accurate picture of the effects of the structural modes on the absorption behavior of the sample. The simulation seems to generally overestimate the stiffness of the melamine foam, which results in a slight shift of the absorption coefficient curve toward higher frequencies, and the absorption is underestimated at higher frequencies. The choice of the material orientations (see left subfigure in Fig. 6) may be one source of errors in the numerical simulation. It is worth noting that the absorption coefficients measured reflect the acoustical behavior of the wedges, the wooden frame, and the remaining air gap and not the absorption coefficient of the acoustic lining actually installed in the anechoic chamber. The lower right subfigure in Fig. 7 gives the numerical results obtained on the configuration consisting of a 3×3 sample of wedges placed in an impedance tube measuring 0.9×0.9 m^2, which corresponds to the properties of the melamine wedges installed in the anechoic chamber. The sample of acoustic lining was found to have an acceptable absorption coefficient in the 50–75 Hz frequency range, which is in good agreement with the experimental results presented in Sec. II. A significant decrease in the absorption was found to occur in the 110–180 Hz frequency range. This behavior of the melamine wedges contrasts with the results of the impedance tube experiment described in Sec. II, but is in good agreement with the results of measurements carried out in the anechoic chamber (see Fig. 2), where significant deviations from the free-field conditions were observed in this frequency range. It is worth noting that only when three material axes were used a numerical model was obtained that matched the experimental data. When an isotropic material

TABLE II. Material parameters of melamine identified.

Young's modulus (N/m^2)			Shear modulus (N/m^2)			Poisson's ratio		
E_x	E_y	E_z	G_{xy}	G_{yz}	G_{xz}	ν_{xy}	ν_{xz}	ν_{yz}
0.30*E*6	0.22*E*6	0.50*E*6	0.10*E*6	0.08*E*6	0.21*E*6	0.44	0.26	0.19



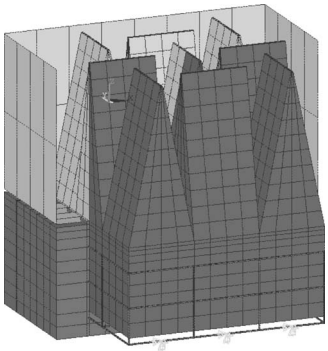


FIG. 6. Photo showing the position of the wedges and the wooden frame in the impedance tube (left subfigure). The numbering of the wedges corresponds to the different material orientations. The meshed geometry of the melamine wedges and surrounding air and (parts of the) walls of the impedance tube (right subfigure).

model and one of Young's moduli given in Table II were used, sharp peaks in the absorption coefficient curve occurred at the frequencies corresponding to structural modes. Figure 8 shows the absorption coefficients calculated and the average longitudinal displacement of the wedge tips. Averaging was carried out with and without respecting the phase, $u_z^1=|\Sigma_i u_{zi}|$ and $u_z^2=\Sigma_i|u_{zi}|$, respectively. Up to a frequency of 60 Hz, almost no difference in the absorption coefficient was observed, as the only structural mode occurring at ≈44 Hz was not affected by the stiffness of the foam. This mode was mainly affected by the stiffness of the iron frame. However, the absorption peak occurring at 85 Hz at $E_z=0.22\times10^6$ N/m^2 shifted to 95 Hz at $E_z=0.3\times10^6$ N/m^2 and to 110 Hz at $E_z=0.5\times10^6$ N/m^2, which shows the existence of a link between Young's modulus of the material and its absorption coefficient. A similar shift was observed, although to a lesser extent, in the case of the peak at 65 Hz at $E_z=0.22\times10^6$ N/m^2 which occurred at 70 and 90 Hz with the other material orientations. Upon analyzing the average longitudinal displacement u_z^2 of the wedges, the structural modes were found to have shifted in parallel with the absorption coefficient. Higher absorption values were observed only when the wedges vibrated out of phase. When the wedges were vibrating in phase, i.e., with peaks in u_z^1 and u_z^2, no additional absorption occurred. The effects of the structural modes of the wedges on the absorption of the lining show that the real mounting conditions must be respected when carrying out impedance tube experiments, as these conditions significantly affect the structural eigenfrequencies and modes.

To model the absorption coefficients measured, which showed less strongly expressed peaks (compare Figs. 7 and 8), at least three different material axes had to be considered, as shown in the lower right subfigure in Fig. 8. Spreading out the structural modes of the wedges over the 110–170 Hz frequency range may reduce the decrease in the absorption observed in the present configuration (see lower right subfig-

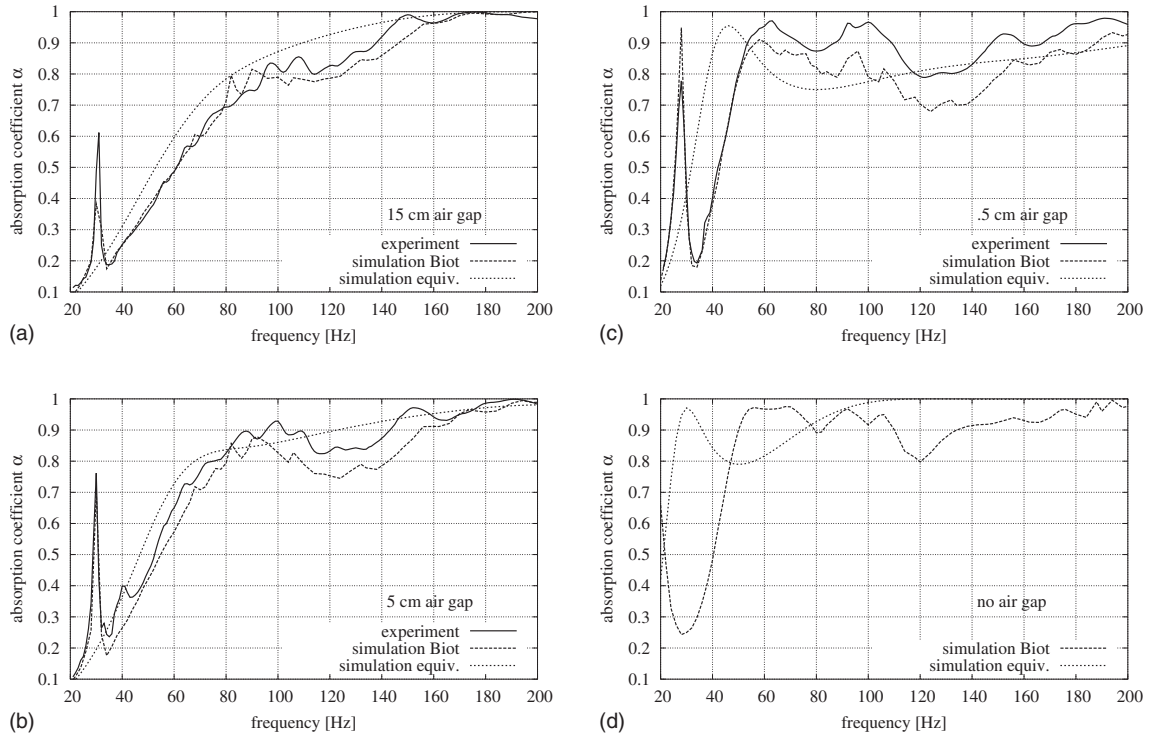


FIG. 7. Comparison between the measured and calculated absorption coefficients of the 3×3 sample of wedges with various air gap sizes. The Biot model and the equivalent fluid model were used to simulate the behavior of the absorbing material numerically.

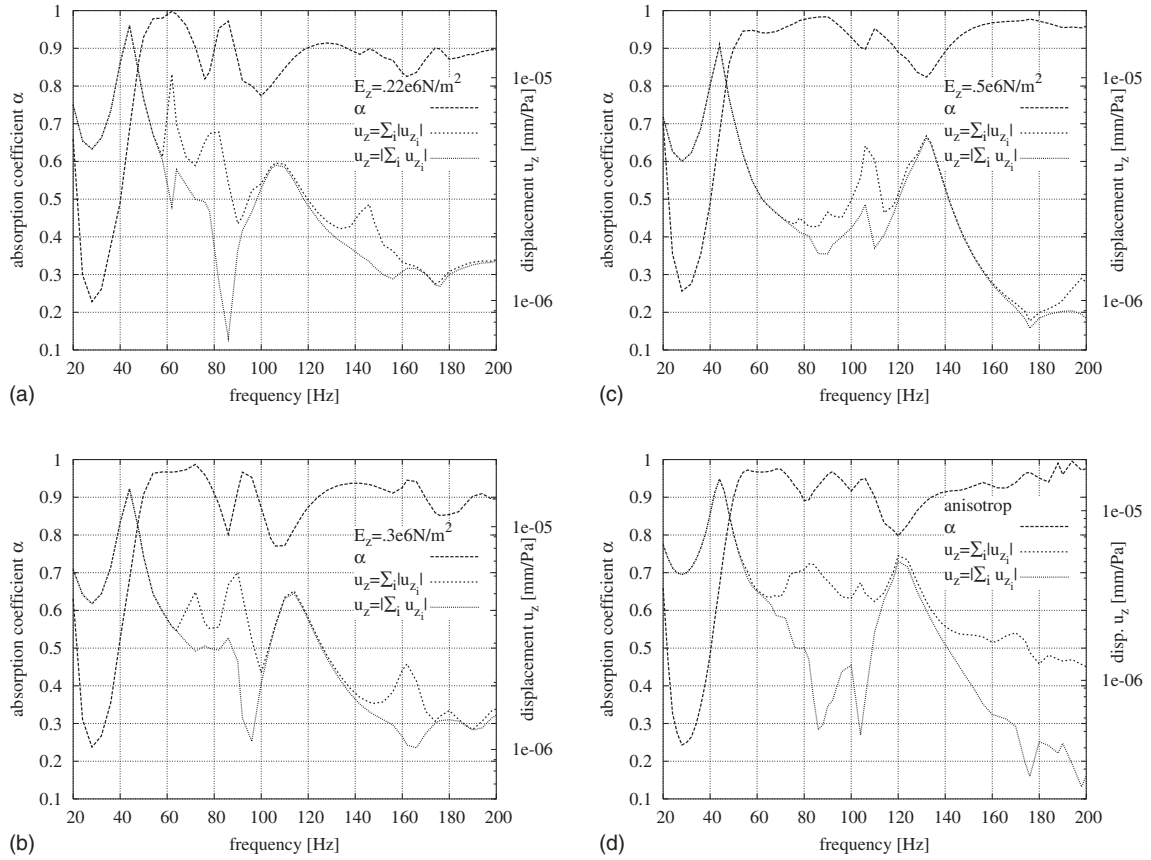


FIG. 8. Influence of the stiffness of the elastic frame on the absorption coefficient and the longitudinal displacement of the wedges.

ure in Fig. 8). This strategy was already partially used to advantage when the acoustic lining was installed by giving the wedges a random orientation relative to the material axes. Varying the length of the wedges in addition to the material axes provides a further means of obtaining a lining with a less fluctuating absorption coefficient. To test this hypothesis, for each material orientation three wedges with different tapering section lengths, 0.6, 0.7, and 0.8 m, were tested in a numerical study. The result of a numerical simulation on a sample of the acoustic lining of this kind are compared with the data obtained on the real lining in Fig. 8. The absorption coefficient of the sample consisting of wedges of different lengths showed fewer variations and a much less severe breakdown at around 120 Hz. It is therefore possible, with almost the same quantity of absorbing material to obtain a more efficient acoustic lining. (Fig. 9)

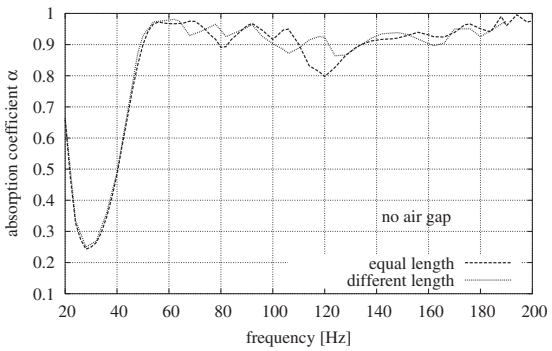


FIG. 9. Influence of the length of the wedges on the absorption coefficient.

It is worth noting that assuming the frame of the melamine foam to be motionless, i.e., using the equivalent fluid model instead of the complete Biot model, yields absorption coefficients (see Fig. 7) that do not reflect the actual behavior of the melamine wedges, as their acoustical behavior is determined by the elastic frame.

V. CONCLUSIONS

In this study, two different methods of measuring the acoustic properties of the acoustic lining of the anechoic chamber at the LMA were compared. The experimental results obtained showed that the mounting conditions of the acoustic lining significantly affect the absorption coefficient, and this factor must therefore be taken into account when assessing the quality of the lining. Structural modes of the wedges were found to increase the absorption at the corresponding frequencies. Modifying the eigenfrequencies of these modes is one possible means of optimizing the properties of the acoustic lining. Spreading out the modes where the wedges vibrate with a phase shift over the frequency range where higher absorption is required can help to improve the performance of a lining. The values of the eigenfrequencies can be controlled by using various materials or various orientations of an anisotropic material and by using wedges of various lengths.

Numerical simulations showed that the anisotropy of the melamine foam must be taken into account to obtain accurate results. In addition, assuming the frame of the melamine foam to be motionless leads to unacceptable errors. Since the

validity of the numerical model for the acoustic lining developed here has been confirmed, it will be used in an ongoing project to simulate an entire anechoic chamber.

ACKNOWLEDGMENTS

The author was financed by a grant from the "German Research Foundation" and by the French "Agence Nationale de le Recherche" under Project No. BLAN06-134753. Numerical simulations were run on an SGI Origin3800 at the Center for Information Services and High Performance Computing at the Technische Universität Dresden, Germany.

[1]L. L. Beranek and J. H. P. Sleeper, "The design of anechoic sound chambers," J. Acoust. Soc. Am. **18**(1), 246–246 (1946).
[2]K. A. Cunefare, V. B. Biesel, J. Tran, R. Rye, A. Graf, M. Holdhusen, and A.-M. Albanese, "Anechoic chamber qualification: Traverse method, inverse square law analysis method, and nature of test signal," J. Acoust. Soc. Am. **113**, 881–892 (2003).
[3]H. C. Hardy, F. G. Tyzzer, and H. H. Hall, "Performance of the anechoic room of the parmly sound laboratory," J. Acoust. Soc. Am. **19**, 992–995 (1947).
[4]G. R. Hruska and W. Koidan, "Evaluation of the large anechoic chamber at the national bureau of standards," J. Acoust. Soc. Am. **58**, S78–S78 (1975).
[5]W. Koidan and G. R. Hruska, "Acoustical properties of the national bureau of standards anechoic chamber," J. Acoust. Soc. Am. **64**, 508–516 (1978).
[6]J.-Q. Wang and B. Cai, "Calculation of free-field deviation in an anechoic room," J. Acoust. Soc. Am. **85**, 1206–1212 (1989).
[7]W. Koidan, G. R. Hruska, and M. A. Pickett, "Wedge design for national bureau of standards anechoic chamber," J. Acoust. Soc. Am. **52**, 1071–1076 (1972).
[8]G. E. Warnaka, "A different look at the anechoic wedge," J. Acoust. Soc. Am. **75**, 855–858 (1984).
[9]J. Xu, J. M. Buchholz, and F. R. Fricke, "Flat-walled multilayered anechoic linings: Optimization and application," J. Acoust. Soc. Am. **118**, 3104–3109 (2005).
[10]X. Olny and C. Boutin, "Acoustic wave propagation in double porosity media," J. Acoust. Soc. Am. **114**, 73–89 (2003).
[11]F. Sgard, X. Olny, N. Atalla, and F. Castel, "On the use of perforations to improve the sound absorption of porous materials," Appl. Acoust. **66**, 625–651 (2005).
[12]V. Easwaran and M. L. Munjal, "Finite element analysis of wedges used in anechoic chambers," J. Sound Vib. **160**, 333–350 (1993).
[13]Y. J. Kang and J. S. Bolton, "Finite element models for sound transmission through foam wedges and foam layers having spatially graded properties," J. Acoust. Soc. Am. **98**, 2976–2976 (1995).
[14]Y. J. Kang and J. S. Bolton, "Sound transmission through elastic porous wedges and foam layers having spatially graded properties," J. Acoust. Soc. Am. **102**, 3319–3332 (1997).
[15]C.-N. Wang and M.-K. Tang, "Boundary element evaluation on the performance of sound absorbing wedges for anechoic chambers," Eng. Anal. Boundary Elem. **18**, 103–110 (1996).
[16]T. Kar and M. Munjal, "Plane wave analysis of acoustic wedges using the boundary-condition-transfer algorithm," Appl. Acoust. **67**, 901–917 (2006).
[17]D. Pilon, R. Panneton, and F. Sgard, "Behavioral criterion quantifying the edge-constrained effects on foams in the standing wave tube," J. Acoust. Soc. Am. **114**, 1980–1987 (2003).
[18]ISO 10534-2:1998 Acoustics-Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method.
[19]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. i. theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[20]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. ii. experiment," J. Acoust. Soc. Am. **68**, 914–921 (1980).
[21]J.-F. Allard, *Propagation of Sound in Porous Media. Modelling Sound Absorbing Materials* (Elsvier, London, 1993).
[22]M. Melon, C. Langrenne, D. Rousseau, B. Roux, and P. Herzog, "Comparison of four subwoofer measurement techniques," In *120th AES Convention*, number 6833, Paris (France), 2006.
[23]S. Schneider and C. Kern, "Acoustical behavior of the large anechoic chamber at the Laboratoire de Mécanique et d'Acoustique in the low frequency range," Acust. Acta Acust. **94**, 141–147 (2008).
[24]N. Dauchez, "Étude vibroacoustique des matériaux poreux par éléments finis (Vibroacoustic analysis of poroelastic materials using the finite element method)," Ph.D. thesis, Université de Sherbrooke, 1999.
[25]D. Pilon and R. Panneton, "Effects of circumferential air gaps on the measurement of the absorption coefficient of poroelastic materials," J. Acoust. Soc. Am. **112**, 2383–2383 (2002).
[26]B. H. Song, J. S. Bolton, and Y. J. Kang, "Effect of circumferential edge constraint on the acoustical properties of glass fiber materials," J. Acoust. Soc. Am. **110**, 2902–2916 (2001).
[27]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[28]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[29]N. Atalla, R. Panneton, and P. Deberge, "A mixed pressure-displacement formulation for poroelastic materials," J. Acoust. Soc. Am. **104**, 1444–1542 (1998).
[30]J. Rahola, "Diagonal forms of the translation operators in the fast multipole algorithm for scattering problems," BIT **60**, 333–358 (1996).
[31]V. Rokhlin, "Diagonal forms of the translation operators for the Helmholtz equation in three dimensions," Appl. Comput. Harmon. Anal. **1**, 82–93 (1993).
[32]W. C. Chew, J. M. Jin, C. C. Lu, E. Michelssen, and J. M. Song, "Fast solution methods in electromagnetics," IEEE Trans. Antennas Propag. **45**, 533–543 (1997).
[33]E. Darve, "The fast multipole method I: Error analysis and asymptotic complexity," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **38**, 98–128 (2001).
[34]J. Song, C.-C. Lu, and W. C. Chew, "Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects," IEEE Trans. Antennas Propag. **45**, 1488–1493 (1997).
[35]S. Schneider, "FE/FMBE-coupling to model fluid-structure interaction," Int. J. Numer. Methods Eng. (to be published).
[36]L. Jaouen, "Contribution à la caractérisation mécanique de matériaux poro-visco-élastiques en vibro-acoustique (Contribution to the mechanical characterization of poro-visco-elastic materials in virbo-acoustics)," PhD thesis, Université de Sherbrooke, 2003.
[37]M. Melon, E. Mariez, C. Ayrault, and S. Sahraoui, "Acoustical and mechanical characterization of anisotropic open-cell foams," J. Acoust. Soc. Am. **104**, 2622–2627 (1998).

A single measurement approach for the determination of the normal incidence transmission loss

Paolo Bonfiglio and Francesco Pompoli
Dipartimento di Ingegneria, Università degli Studi di Ferrara, Via Saragat 1, 44100 Ferrara, Italy

(Received 4 December 2007; revised 4 June 2008; accepted 4 June 2008)

Several techniques for the determination of the normal incidence complex transmission coefficient and transmission loss have been proposed in literature. It has been shown that two different measurements have to be carried out for a correct evaluation of the above-mentioned quantities. However, single measurement approaches can be used at the condition that energy contributions are correctly included in the implemented formulation. This paper presents a single measurement approach based on a transfer matrix formulation, taking into account reflection contribution from the end termination and phase shift introduced by the material. The transmission performances of different kinds of layered porous materials are investigated by means of a four microphone technique in a plane wave tube. By using the decomposition technique, incident, reflected, and transmitted contributions are separated and transmission coefficient is easily calculated. Results are discussed and compared with similar techniques and with the two-load method.
 [DOI: 10.1121/1.2951605]

PACS number(s): 43.55.Rg, 43.20.Ye [NX] Pages: 1577–1583

I. INTRODUCTION

In different areas of acoustics there is the need to experimentally determine the capability of systems to isolate sound waves propagating through them. As an example, multilayer systems made of both rigid or porous materials are widely used in the automotive field, and the knowledge of the transmission coefficient is useful for manufacturers and companies in the choice of an adequate isolating package or for the prediction of the sound field in the cabin by means of numerical codes.

Despite the use of standards[1,2] for measuring the diffuse field transmission loss of layered structures, normal incidence measurements in an impedance tube are demonstrated to be able to provide reliable results even if small samples are tested.

At present several experimental techniques have been developed to test normal incidence acoustical performance of both homogeneous and layered systems. All these approaches are based on the decomposition of the plane wave field into incident, reflected, and transmitted waves.

It is worth emphasizing that in the case of nonsymmetric systems, it can be proven that a complete solution for the normal incidence reflection and transmission problem requires two measurements, for example, by using different sources or end terminations.[3,4]

In literature different methods requiring a single measurement for evaluating transmission properties of materials can be found. One of the first techniques was proposed by Chung and Blaser,[5] and it was widely used in the past. The limitation of this method is the assumption of a perfectly anechoic termination. Afterward, Song and Bolton[6] proposed a transfer matrix technique for measuring transmission coefficients for homogeneous and isotropic materials. Recently Liu *et al.*[7] proposed a revision of Chung and Blaser's formula, taking into account the reflection from the end termination. Several tests were carried out by Pispola and Horoshenkov[8] in order to compare all these experimental procedures.

In this paper an approach for determining the complex transmission coefficient and transmission loss of layered structures is presented. It originates from the transfer matrix technique presented in Ref. 6, rewritten taking into account the reflection contribution from the not perfectly anechoic termination.

This paper is organized as follows. The description of the methods proposed in literature together with the revised formulation for the determining of the complex transmission coefficient and transmission loss will be presented in the next section. In the third section the measurement setup and procedure will be discussed, and the fourth section will report the results for tested homogeneous and layered porous systems; moreover a comparison between different terminations and an estimation of the measurement uncertainty will be provided. Finally, concluding remarks will be given in the last section.

II. THEORETICAL BACKGROUND

By considering the measurement setup shown in Fig. 1 and by using the decomposition technique, the contributions of the incident and reflected acoustical waves at both sides of the material can be calculated by measuring complex pressures at the different microphone positions as follows:

$$A = \frac{j(P_1 e^{jk_0 x_2} - P_2 e^{jk_0 x_1})}{2\sin[k_0(x_1 - x_2)]}, \tag{1}$$

$$B = \frac{j(P_2 e^{-jk_0 x_1} - P_1 e^{-jk_0 x_2})}{2\sin[k_0(x_1 - x_2)]}, \tag{2}$$

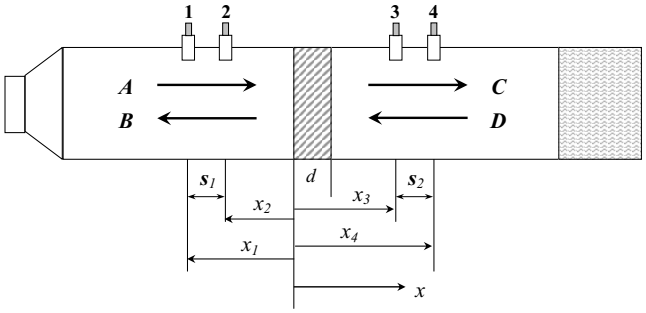


FIG. 1. Measurement setup scheme.

$$C = \frac{j(P_3 e^{jk_0 x_4} - P_4 e^{jk_0 x_3})}{2\sin[k_0(x_3 - x_4)]}, \tag{3}$$

$$D = \frac{j(P_4 e^{-jk_0 x_3} - P_3 e^{-jk_0 x_4})}{2\sin[k_0(x_3 - x_4)]}, \tag{4}$$

where P_i are the sound pressures at the four microphone positions, here indicated as x_1–x_4, and k_0 is the wave number. It has to be mentioned that wave number k_0 could be complex valued to take into account the effects of viscous and thermal dissipation in the oscillatory thermoviscous boundary layer that forms on the inner surface of the duct. This work has chosen to use an analytical correction for viscous effects and thermal exchanges in a circular section tube, worked out by Zwikker and Kosten[9] and Stinton.[10]

A. Methods proposed in literature

Now the complex amplitudes (1)–(4) can be related in a matrix formulation as follows:

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \tau & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix}. \tag{5}$$

Equation (5) is a general formulation, and the elements of the matrix depend only on the physical properties of the system. For instance the complex transmission coefficient is proven to be equal to τ^{-1}.

The complete solution of system (5) requires two different measurements using different end terminations. In such conditions it can be proven that

$$T_{2\ \mathrm{load}} = \frac{C^{\mathrm{I}} D^{\mathrm{II}} - C^{\mathrm{II}} D^{\mathrm{I}}}{A^{\mathrm{I}} D^{\mathrm{II}} - A^{\mathrm{II}} D^{\mathrm{I}}}, \tag{6}$$

where the symbols I and II indicate the measures corresponding to the different terminations.

According to the Chung and Blaser[5] approach, it $D=0$ should be imposed simply and then the complex transmission coefficient can be calculated by means of a single measurement as follows:

$$T_{C-B} = \frac{C}{A}. \tag{7}$$

In reality, it is extremely difficult to create a perfectly absorbing termination; hence a correction of the Eq. (7) was necessary in order to evaluate the correct propagation of the waves in the duct. Therefore Liu *et al.*[7] provided a revised formula for the normal incidence transmission coefficient, that is,

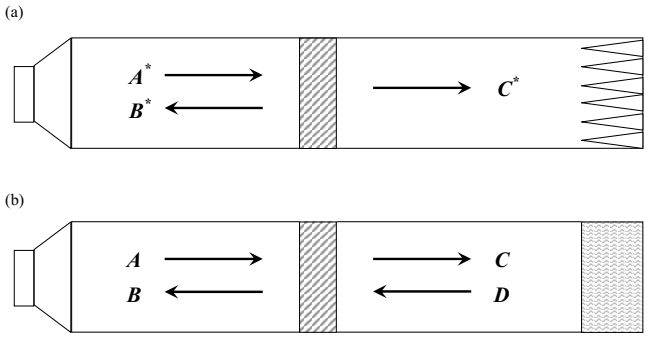


FIG. 2. Measurement tube in two different configurations: (a) tube with anechoic termination; (b) tube with reflecting termination.

$$T_{\mathrm{Liu}} = \frac{AC - BD}{A^2 - D^2}. \tag{8}$$

It must be kept in mind that this method presumes the validity of the conditions of symmetry and reciprocity; that is to say, reflection and transmission coefficients are assumed to be the same at both directions of incidence.[11,12] Reciprocity condition for attenuating layered materials has been proven by Allard *et al.*[13]

A different transfer matrix formulation was proposed by Song and Bolton,[6] which related sound pressure P and particle velocity V at both surfaces of the tested sample by means of the following expression:

$$\begin{pmatrix} P \\ V \end{pmatrix}_{x=0} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} P \\ V \end{pmatrix}_{x=d}. \tag{9}$$

Even if their work was mainly concerned with the determination of the complex characteristic properties of homogeneous and isotropic materials, they proposed some formulations based on transfer matrix elements for calculating the transmission coefficient of these systems. Starting from Eq. (9) the authors proved that the complex transmission coefficient can be written in the following form:

$$T_{S-B} = \frac{2e^{ik_0 d}}{T_{11} + (T_{12}/\rho_0 c_0) + (T_{21}\rho_0 c_0) + T_{22}}, \tag{10}$$

where ρ_0 (kg/m^3) is the ambient fluid density and c_0 (m/s) is the ambient sound speed, and d (m) is the thickness of the material.

It is interesting to notice that it is not possible to determine each of the matrix elements T_{ij}, and two additional equations are required. An indirect way to obtain these relations is to repeat measurements with two different end terminations. It can be proven that results obtained using this approach are identical to those obtained using the standard two-load method [Eq. (6)].

B. Revised formulation

A qualitative interpretation of the revised formulation will be given in the following paragraph.

In Fig. 2 measurement tube is shown in two different configurations: (a) tube with perfectly anechoic termination

and (b) tube with a reflecting termination. In the former case the ratio between terms B^* and A^* corresponds to the true reflection coefficient (R^*) of the layer with a semi-infinite air space on both sides. Similarly the ratio between terms C^* and A^* is the true transmission coefficient (T^*) when no component from the end termination contributes at the total plane wave sound field within the tube.

In the latter case the contribution from the termination cannot be excluded; consequently it is possible to write the term B as the sum between the true reflected contribution (B^*) and the transmitted contribution from the end termination (T^*D). Similarly, the measured term C can be thought of as the sum of the true transmitted contribution (C^*) and the end termination contribution reflected from the material (R^*D). It has to be emphasized that proposed formulation assumes the validity of symmetry condition.

Now, by using the transfer matrix approach [Eq. (9)], it is possible to write the elements T_{ij}, which are functions of pressures and particle velocities at both sides of the tested sample, in terms of the measured incident (A), reflected (B), transmitted (C), and reflected from the termination (D) components. The relations between pressures and particle velocities at both sides of the layer and the plane wave propagating contributions can be expressed as follows:

$$P_{x=0} = A + B,$$

$$V_{x=0} = \frac{A - B}{\rho_0 c_0},$$

$$P_{x=d} = Ce^{-jk_0 d} + De^{jk_0 d},$$

$$V_{x=d} = \frac{Ce^{-jk_0 d} - De^{jk_0 d}}{\rho_0 c_0}. \tag{11}$$

After simple manipulations, it is possible to obtain a revised formula both for effective complex transmission and reflection coefficients by solving the following system:

$$R = R^* + T^* R_1 T,$$

$$T = T^* + R^* R_1 T e^{2jk_0 d}, \tag{12}$$

where $R=B/A$, $T=C/A$, and $R_1=D/C$. Here the exponential term takes into account the phase difference between pressures and particle velocities at both sides of the tested sample.

It can easily be proven that complex transmission coefficient can be calculated, with a single measurement, using following formula:

$$T^* = \frac{T(1 - RR_1 e^{2jk_0 d})}{1 - (R_1 T)^2 e^{2jk_0 d}}. \tag{13}$$

Of course it is important to remember that Eq. (13) is rigorously valid for symmetric systems. Later it will be demonstrated that it is more accurate than Eqs. (7) and (8) if compared with the two-load method [Eq. (6)] or equivalently with Eq. (10).

The main differences between the proposed formulation and Chung and Blaser and Liu methods are due to the diverse physical interpretation of the combination of wave contributions in the sound field. In fact, as already mentioned in Sec. II A, the approach proposed by Chung and Blaser assumes that no contribution comes from the end termination (that is, $D=0$ in Fig. 1); in practical situations it is difficult to realize a perfectly anechoic termination. Thus the aforementioned approximation is not correct. The formulation of Liu *et al.* takes into account the contribution from the backing, but the term including the phase shift introduced by the system is not considered. On the contrary, these two different effects are included in the proposed approach.

In all the cases examined the transmission loss is calculated from the complex transmission coefficient as

$$\mathrm{TL} = 20 \log_{10}\left(\frac{1}{|T_p|}\right), \tag{14}$$

where T_p is the generic complex transmission coefficient.

III. MEASUREMENT SETUP AND PROCEDURE

The test setup consists of a plane wave tube with a diameter of 0.045 m, equipped with four microphone positions and the material sample under test placed in the middle. On one end there is a loudspeaker, and the other end consists of a 0.25 m sample of polyester fiber material (25 kg/m^3) placed on a rigid backing.

The distance between microphones is equal to 0.03 m, and it can be extended up to 0.1 m. Thus, following the same rules as for the transfer function method,[14] the working frequency range is fixed between 50 and 4300 Hz.

The measurement chain consists of a PC equipped with the National Instruments PCI-6120 for data acquisition, four $\frac{1}{4}$ in. condenser microphones B&K 4939, signal conditioner MESA MUX10A, and power amplifier B&K 2716C. The entire measurement procedure was implemented in the form of a homemade graphical interface, developed on the LABVIEW platform.

The sound pressures at the four positions are measured by impulse responses, obtained with the exponential sine sweep method.[15] Because of the transient characteristic of the impulse response signals, it is enough to perform a single fast Fourier transform operation (i.e., without averages and overlapping procedures) to get the complete amplitude and phase information, in frequency domain, at each microphone position.

The amplitude and phase calibration procedures of the four microphones are obtained by a series of cross measurements, without any test sample, at two predetermined positions (e.g., positions 3 and 4 in Fig. 1), respectively. It is then possible to define four complex calibration pressures, one for each microphone, as follows:

$$P_{ic} = \sqrt{P_{iL} P_{iR}}, \quad i = 1,2,3, \text{ and } 4. \tag{15}$$

In order to determine the calibrated pressures, simply divide measured pressures by the relative calibration pressure.

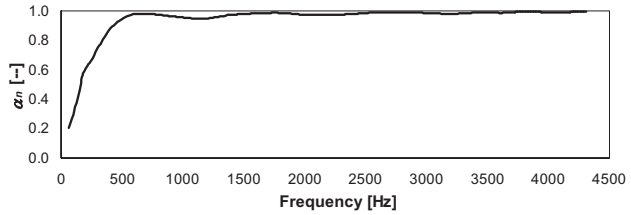


FIG. 3. Normal incidence sound absorption coefficient of the end termination.

IV. RESULTS AND DISCUSSIONS

A. Comparison between different methods

Different porous structures, both homogeneous and layered, were investigated, and the diverse methods for measuring complex transmission coefficients presented in Sec. II were compared.

As previously mentioned, all the methods for determining the transmission loss provide the same results only if a perfectly anechoic termination is used. Otherwise, it is necessary to take into account the reflected wave component from the termination. Before presenting experimental results, it is interesting to analyze the sound absorption coefficient of the end termination, depicted in Fig. 3, that was used for the tests.

From the figure it can be noticed that the sound absorption coefficient is lower than 1 almost over all the frequency range, and thus the termination cannot be considered perfectly anechoic.

It is important to underline that a rigid backing was used for the additional measurement required by using the two-load method. In Table I, a description of the porous materials used for the tests is reported. In Fig. 4 curves of transmission loss are reported for a homogeneous fibrous material manufactured from polyester. In order to enhance the clarity of the figures, the two-load method is firstly compared to Chung and Blaser and Liu approaches, and then the comparison between the two-load method and revised formulation results is shown in a different graph.

It is seen that the results of the approach proposed here are good if compared with the two-load method. By contrast, the Chung and Blaser and Liu results show some differences across the entire frequency range since the end termination is not perfectly anechoic.

It has to be strongly emphasized that since a homogeneous material was used, all the results should have been similar because of the validity of the symmetry condition in this particular case.

TABLE I. A summary of the investigated porous materials used for the transmission loss tests.

Material	Symbol	Density (kg/m^3)	Thickness (mm)	Typology
Polyester fiber	M1	40	10	Fibrous
Reticulated foam	M2	10	20	Porous
Consolidated flint	M3	1500	30	Cellular
Reconstituted porous rubber	M4	64	24	Cellular

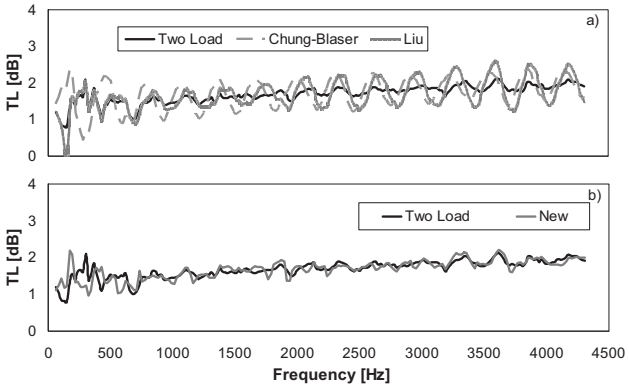


FIG. 4. Transmission loss for a homogeneous material (M1). (a) Comparison between the two load method and Chung–Blaser and Liu approaches. (b) Comparison between the two load method and new formulation results.

In Fig. 5 the results for a symmetric layered structure are shown. The system consists of reticulated foam, consolidated flint, and again reticulated foam. In Fig. 6 the same comparison is reported for an asymmetric system obtained by coupling the reconstituted porous rubber and the previous layered structure. In both cases it is noticed that the results obtained using the proposed formulation are closer to the two-load formulation than those obtained by using Eqs. (7) and (8). Less marked differences are seen for frequencies higher than around 2000 Hz; on the contrary at lower frequencies the reflections (in terms of energy and phase) from the end termination are significant.

Analyzing the results it is interesting to note that results obtained using the Chung and Blaser and Liu approaches may present different uncertainties even though the same end termination is used. In fact in the case of the homogeneous material, since the value of the transmission loss is lower than that other system of materials, it is possible to observe deviation up to 2 dB in the entire frequency range. In the cases of symmetrical and asymmetrical systems, deviation lies between 3 and 6 dB for frequencies lower than 2000 Hz and reduces appreciably in the frequency range between 2000 and 4300 Hz. In all the cases the difference between the two-load method and the proposed formulation is lower than 2 dB.

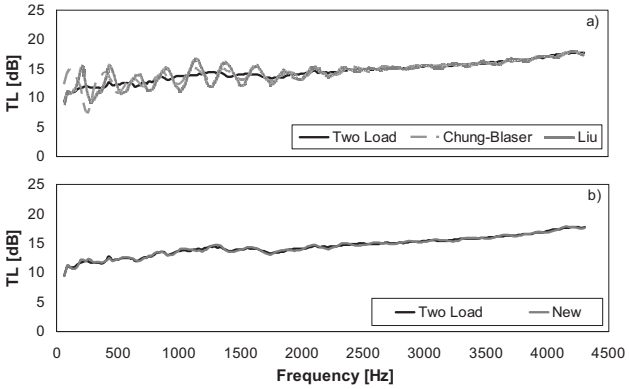


FIG. 5. Transmission loss for a symmetrically layered material (M2+ M3 + M2). (a) Comparison between the two load method and Chung–Blaser and Liu approaches. (b) Comparison between the two load method and new formulation results.

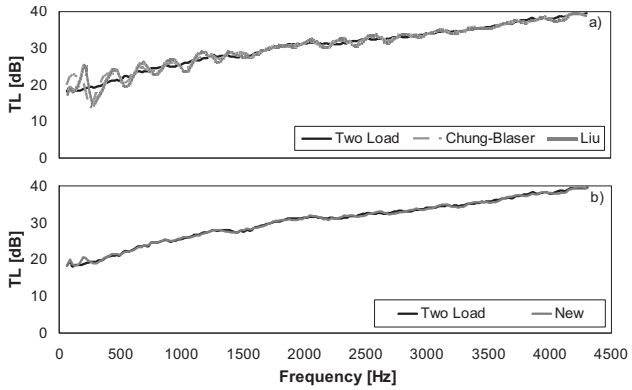


FIG. 6. Transmission loss for an asymmetrically layered material (M2+M3+M2+M4). (a) Comparison between the two load method and Chung–Blaser and Liu approaches. (b) Comparison between the two load method and new formulation results.

From the analysis it appears that the proposed formulation better accounts for the effect of the wave contribution from the backing. To this end in Fig. 7 the comparison of the amplitude and angle of the complex transmission coefficient related to the tested structures mentioned above is reported. It can be observed that the proposed formula is reliable for evaluating the transmission performance of the examined systems. In fact the amplitude difference between the two-load method and new formulation does fall below 0.08 for frequencies higher than 500 Hz and is lower than 0.05 in the frequency range of 50–500 Hz. Regarding the phase difference, it is not greater than 0.15 rad in the entire frequency range.

In order to explain the capability to measure the transmission performances of asymmetric systems by means of the proposed formulation, it has to be kept in mind that a highly absorbing termination was used during tests; thus the effect of the end contribution on system (12) is strongly reduced even if reflection coefficients of the tested system at both directions are different. In next paragraph it will be shown that when a different end termination is used, some deviations can be observed with respect to the two-load method; anyway such differences are small if compared to the ones related to formulations (7) and (8) when the absorbing termination was used.

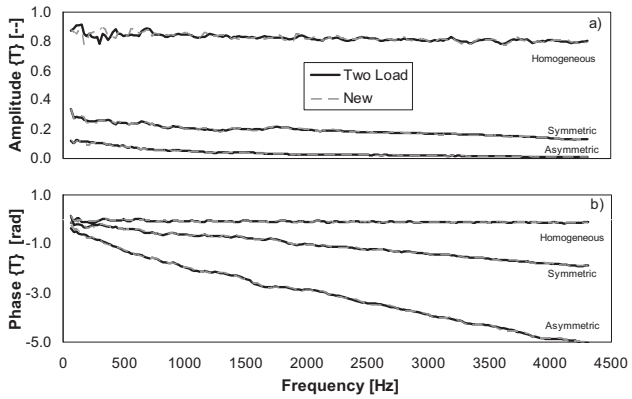


FIG. 7. Comparison between the two load method and new formulation results for the tested structures. (a) Amplitude of the complex transmission coefficient. (b) Phase of the complex transmission coefficient.

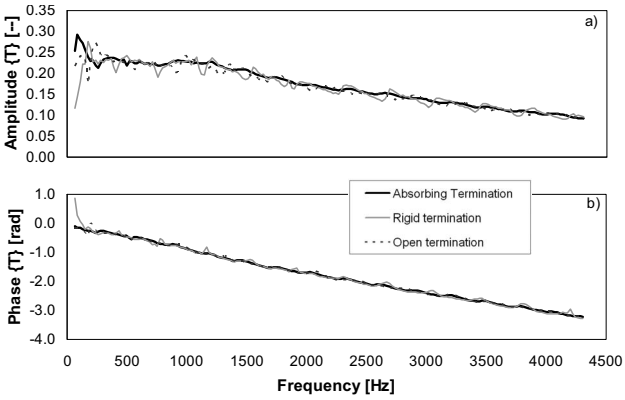


FIG. 8. Comparison between different end terminations (absorbing, rigid, and open) for the material M4. (a) Amplitude of the complex transmission coefficient. (b) Phase of the complex transmission coefficient.

B. Comparison between different end terminations

The effect of different end termination on the results from the revised formulation has been investigated. For this test the absorbing termination, as described in Sec. III, the rigid termination, and the open tube termination have been used, and curves of the amplitude and angle of the complex transmission coefficient are shown in Fig. 8. From the figure it can be noticed that no appreciable differences can be identified, except that some peaks occur in the curves obtained using the rigid and open terminations. Analyzing the results, in all the cases it can be underlined that the amplitude differences are lower than 0.1 and the phase difference is lower than 0.3 rad.

C. Measurement uncertainty estimation

Several researches have pointed out that a complete analysis of measurement uncertainties in measuring acoustical properties of materials put in a plane wave tube is extremely difficult to be carried out.[16] In particular, different effects (such as the compression of the sample within the tube, lateral leakages, and constrains) can affect the final results and rarely it is possible to account properly for them in the error analysis.

For such reason, in this section a simplified formulation for estimating errors associated only with mathematical formulation of the proposed method is presented. In fact, analyzing Eqs. (13) and (14) and neglecting uncertainties due to the determination of microphone positions and thicknesses of the sample, it is possible to notice that only variations of complex sound pressures could lead to errors in the determination of the complex transmission coefficient.

The error propagation rules for complex functions have been used for the determination of the errors on amplitude and phase of the complex transmission coefficient and transmission loss as a function of the real and imaginary parts of the sound pressures together with their uncertainties. It has to be emphasized that variables are not independent and the method of moments[17] (involving the use of the correlation relations between errors) needs to be used for a correct evaluation of the measurement errors. According to that method it is possible to write the same expression for calcu-

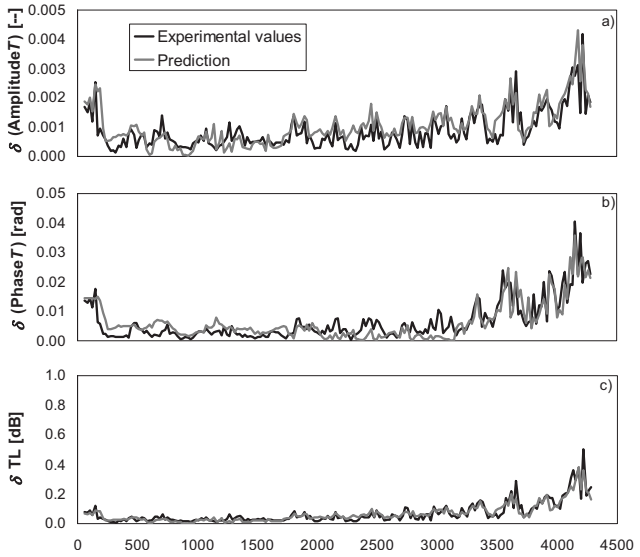


FIG. 9. Comparison between experimental and predicted uncertainties for the material M4. (a) Error on the amplitude of the complex transmission coefficient. (b) Error on the phase of the complex transmission coefficient. (c) Error on the transmission loss.

lating errors on the amplitude and phase of the complex transmission coefficient and the transmission loss,

$$\delta(\Psi) = \sum_{i=1}^{4} \left[\left(\frac{\partial \Psi}{\partial \mathrm{Re}\{P_i\}} \delta \mathrm{Re}\{P_i\} \right)^2 + \left(\frac{\partial \Psi}{\partial \mathrm{Im}\{P_i\}} \delta \mathrm{Im}\{P_i\} \right)^2 + 2r \left(\frac{\partial \Psi}{\partial \mathrm{Re}\{P_i\}} \delta \mathrm{Re}\{P_i\} \right) \times \left(\frac{\partial \Psi}{\partial \mathrm{Im}\{P_i\}} \delta \mathrm{Im}\{P_i\} \right) \right]^{1/2}, \quad (16)$$

where $\mathrm{Re}\{P_i\}$, $\mathrm{Im}\{P_i\}$, $\delta \mathrm{Re}\{P_i\}$, and $\delta \mathrm{Im}\{P_i\}$ are the real and imaginary parts of the complex sound pressures and their absolute errors, and r is the correlation coefficient between the real and imaginary parts of each sound pressure. In the previous equation the symbol Ψ has been adopted to generically indicate the amplitude and phase of the complex transmission coefficient and the transmission loss.

For the validation of Eq. (16) material M4 was tested and sound pressures were measured five times at each microphone position. In order to neglect the effects of the mounting of the sample, measurements were repeated without moving the specimen. In this case the main error sources were the measurement chain and the signal analysis. The errors $\delta \mathrm{Re}\{P_i\}$ and $\delta \mathrm{Im}\{P_i\}$ were calculated from the standard deviation of the measurements. In terms of amplitude and phase, the uncertainties were lower that 0.02 Pa and 0.9 rad.

For the sake of simplicity, errors have been supposed to be perfectly correlated ($r=1$). Although this assumption could lead to an overestimation of the errors, it certainly simplifies the computation of Eq. (15). Figure 9 shows the comparison between errors on the amplitude and phase of the complex transmission coefficient and the transmission loss determined from the experimental data and the ones predicted by means of Eq. (15) in the frequency range between 50 and 4300 Hz.

Analyzing the figure it is possible to notice that the differences between experimental values and predictions are very small; in particular they are lower than 0.002 for the amplitude, 0.012 rad for the phase, and 0.14 dB for the transmission loss.

In conclusion, once the uncertainties on sound pressures at microphone positions are known, Eq. (15) can be used to predict the errors on the amplitude and phase of the transmission coefficient and transmission loss.

V. CONCLUSIONS

In this paper different approaches requiring a single measurement for determining normal incidence complex transmission coefficient and transmission loss of homogeneous and layered structures have been investigated, and results have been compared with the two-load method.

A revised formulation requiring a single measurement has been proposed, and it has been shown that provided results were more reliable if compared with the exact formulation, even when a nonperfectly anechoic end termination is used. From the analysis it can be ended that the reliability of results is strongly dependent on the correct evaluation of the reflected wave from the end termination and phase shift introduced by the material.

Furthermore, different end terminations have been investigated, and it can be concluded that no remarkable differences can be identified, except for some peaks occurring by using the rigid and open terminations.

Regarding the measurement uncertainty, an error propagation approach has been proposed to predict errors on the amplitude and phase of the complex transmission coefficient and the transmission loss as a function of the sound pressures together with their uncertainties. The reliability of the proposed formulation has been validated and confirmed against experimental tests.

It has to be strongly emphasized that although applied to the case of porous materials, the proposed formulation is a general approach and can be used for the determination of the acoustical performance of complex systems such as multitrimmed packages, silencers, and mufflers.[18,19]

ACKNOWLEDGMENTS

This work has been developed within the laboratory of research and technology transfer LAV (Laboratory of Acoustics and Vibration) realized through the contribution of Regione Emilia Romagna–Assessorato Attivita' Produttive, Sviluppo Economico, Piano telematico, PRIITT misura 3.4 azione A-Obiettivo 2.

[1]International Standards Organisation, ISO 140-3: 1995, Acoustics—Measurements of sound insulation in buildings and of buildings elements—Part 3: Laboratory measurements of airborne sound insulation of building elements.
[2]International Standards Organisation, ISO/DIS 15186-1: 2000, *Acoustics—Measurements of sound insulation in buildings and of buildings elements using sound intensity—Part 1: Laboratory conditions.*
[3]M. L. Munjal and A. G. Doige, "Theory of a two source-location method

for direct experimental evaluation of the four-role parameters of an aeroacoustic element," J. Sound Vib. **141** 323–333 (1990).
[4]T. Y. Lung and A. G. Doige, "A time-averaging transient testing method for acoustic properties of piping systems and mufflers," J. Acoust. Soc. Am. **73**, 867–876 (1983).
[5]J. C. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–916 (1980).
[6]B. H. Song and J. S. Bolton, "A transfer-matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials," J. Acoust. Soc. Am. **107** 1131–1152 (2000).
[7]K. Liu, Z. Qijun, and T. Feng "The measurement theory and experiment investigation of transmission loss of material—four transducer method," Proceedings of The 11th International Congress on Sound and Vibration, St. Petersburg, Russia (2004), pp. 1585–1592.
[8]G. Pispola and K. V. Horoshenkov, "Consolidated granular media for sound insulation: Performance evaluation through different methods," Proceedings of 12th International Congress on Sound and Vibration, Lisbon, Portugal (2005).
[9]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[10]M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tunes and generalisation to uniform tunes of arbitrary cross-sectional shape," J. Acoust. Soc. Am. **89**, 550–558 (1991).
[11]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).
[12]J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier, London, 1993).
[13]J. F. Allard, B. Brouard, D. Lafarge, and W. Lauriks, "Reciprocity and antireciprocity in sound transmission through layered materials including elastic and porous media," Wave Motion **17**, 329–335 (1993).
[14]International Standards Organisation, ISO 10534-2: 2001 *Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—Part 2: Transfer-function method*.
[15]A. Farina, "Simultaneous measurement of impulse response and distorsion with a swept-sine technique," 108th Convention of the Audio Engineering Society (2000).
[16]K. Horoshenkov, A. Khan, F.-X. Bécot, L. Jaouen, F. Sgard, A. Renault, N. Amirouche, F. Pompoli, N. Prodi, P. Bonfiglio, G. Pispola, F. Asdrubali, J. Hübelt, N. Atalla, C. K. Amédin, W. Lauriks, and L. Boeckx, "Reproducibility experiments on measuring acoustical properties of rigid-frame porous media (round-robin tests)," J. Acoust. Soc. Am. **122**, 345–353 (2007).
[17]J. R. Taylor, *An Introduction to Error Analysis* (University Science Book, Sausalito, CA, 1982).
[18]P. Bonfiglio, F. Pompoli, and A. T. Peplow, "Normal incidence transmission loss of sandwich structures in a plane wave tube," Proceedings of the 13th International Congress on Sound and Vibration, Vienna, Austria (2006).
[19]P. Bonfiglio, F. Pompoli, and R. Pompoli, "Experimental measurement and FEM prediction of the acoustical performances of mufflers" (in Italian), Proceedings of the 34th Italian Acoustics Association Conference, Florence, Italy (2007).

Experimental verification of a two-sensor acoustic intensity measurement in lossy ducts

Tetsushi Biwa
Department of Mechanical Systems and Design, Tohoku University, Sendai 980-0013, Japan

Yusuke Tashiro and Hiroshi Nomura
Department of Crystalline Materials Science, Nagoya University, Nagoya 464-8603, Japan

Yuki Ueda
Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588, Japan

Taichi Yazaki
Department of Physics, Aichi University of Education, Kariya 448-8542, Japan

(Received 16 January 2008; revised 2 June 2008; accepted 5 June 2008)

Two-sensor method proposed by Fusco *et al.* ["Two-sensor power measurements in lossy ducts," J. Acoust. Soc. Am. **91**, 2229–2235 (1992)] is a novel technique that determines acoustic intensity of a gas column in a wide duct from measurements of pressure based on the boundary layer approximation. For further development of this method, its validity is experimentally tested through comparison with the direct method measuring the pressure and the velocity simultaneously, and its formulation is modified to include the narrow duct range where the duct radius is smaller than the viscous boundary layer thickness of the gas. It is shown that the modified two-sensor method enables quick and accurate evaluation of the acoustic intensity seamlessly from narrow to wide duct ranges. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2953311]

PACS number(s): 43.58.Fm, 43.20.Ye [RR] Pages: 1584–1590

I. INTRODUCTION

Acoustic intensity I represents a time-averaged energy flux accompanied by oscillations of pressure and velocity of a gas. For a sound wave in a duct, the acoustic intensity I is expressed by

$$I = \langle PV \rangle = \tfrac{1}{2}\,\mathrm{Re}[\tilde{p}v], \tag{1}$$

where $P=p(x)e^{i\omega t}$ and $V=v(x)e^{i\omega t}$ are the acoustic pressure and the radial average of axial acoustic particle velocity at position x; the angular brackets $\langle\ \rangle$ represent the time average; Re[ ] and the tilde denote the real part and the complex conjugation, respectively; ω is the angular frequency of the acoustic wave. Whereas the viscous interaction between the acoustic wave and the duct wall decreases I, the thermal interaction can increase I when the duct has a steep temperature gradient.[1,2] Measurement of I plays a key role in deepening the experimental understanding of such acoustic power dissipation and production in ducts.[3–13]

Two experimental methods have been used in the measurements of I. One is the direct method simultaneously measuring P and V using a small pressure sensor and a laser Doppler velocimeter (LDV). This method was first reported in 1998 (Ref. 3) and has been employed for the measurements of the acoustic power of thermoacoustic heat engines[4–8] and the quality factor of an acoustic resonator.[9]

The other method is called a two-sensor method.[10,14–16] In this method, pressures $p_A=p(-\Delta x/2)$ and $p_B=p(\Delta x/2)$ are measured by two pressure sensors separated by a distance Δx, as shown in Fig. 1. The pressure $p(0)$ and the velocity $v(0)$ at the middle $(x=0)$ are theoretically deduced from the sum and the difference of p_A and p_B. The intensity at the middle is then determined from $p(0)$ and $v(0)$ through Eq. (1). In order to accurately estimate $p(0)$ and $v(0)$, Fusco *et al.*[10] proposed to take viscous and thermal interactions of the gas with duct walls into account using the *boundary layer approximation*. Their method has been successfully used in the measurements of acoustic power in thermoacoustic heat engines, particularly those using pressurized gases as the working gas.[11–13]

The two-sensor method has a clear advantage to the direct method, since pressure measurements are much easier and simpler than the LDV that needs installation of seeding particles and use of transparent ducts. However, the intensity estimated from pressures has not been experimentally verified yet by the direct measurements. Besides, the application of the two-sensor method has been limited to *wide ducts* where the duct size, such as the radius r_0 for a circular tube, is much larger than the viscous boundary layer thickness δ formed at the duct wall. Here, δ is given by $\sqrt{2\nu/\omega}$, where ν is the kinematic viscosity of a gas.

The propagation of the acoustic wave in ducts can be classified by the magnitude of the ratio of duct radius r_0 to the viscous boundary layer thickness δ. For example, for 1 bar air δ is 0.22 mm at 100 Hz, and so the ratio of duct radius r_0 to δ easily becomes the order of 100 for tubes with $r_0>22$ mm. The two-sensor method has been found to be useful for such a wide duct. However, the ratio r_0/δ decreases even below unity in the porous media such as the stack of thermoacoustic heat engines, and sound absorbers in architectural acoustics and anechoic chambers. It should be

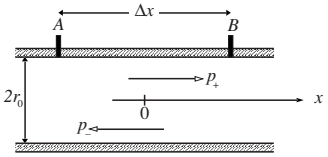


FIG. 1. A cylindrical duct of radius r_0 with two pressure sensors separated by Δx. The origin $x=0$ is taken at the middle of the sensors. p_+ and p_- denote the complex amplitudes of pressure wave traveling in the positive and negative directions.

tested whether the present two-sensor method is applicable to these *narrow ducts*.

In this paper, in order to extract the full ability of the two-sensor method, its validity is experimentally studied by the direct method measuring the pressure and the velocity simultaneously, for ducts with $1.30 \leqslant r_0/\delta \leqslant 82.8$. We show that the two-sensor method becomes inapplicable with $r_0/\delta = 1.3$, and introduce a more general formulation of the method applicable regardless of the ratio r_0 to δ.[17] It is shown that the two-sensor method that we developed enables quick and accurate evaluation of the intensity I in narrow and wide ducts.

II. FORMULATION OF TWO-SENSOR METHOD

We briefly describe here the principle of the two-sensor method referring to the derivation given by Fusco *et al.*[10] Figure 1 schematically explains a cylindrical hollow duct through which the acoustic waves propagate. The locations of two pressure sensors A and B are also shown in Fig. 1. We consider the duct whose radius is much smaller than the wavelength, so the acoustic wave has a plane wave front. Thus, the acoustic pressure is expressed as $P=p(x)e^{i\omega t}$ independent of the radial coordinate, where the x axis is directed from A to B and its origin is taken at the middle of the sensors. We decompose the pressure $p(x)$ as

$$p(x) = (p_+ e^{-ikx} + p_- e^{ikx}), \tag{2}$$

where k is the complex wavenumber. Complex constants p_+ and p_- include the amplitude and the phase of acoustic wave traveling in the positive and negative directions of x. The radial average of the axial velocity $v(x)$ is given from the momentum equation as

$$v(x) = \frac{iF}{\omega\rho}\frac{dp(x)}{dx}, \tag{3}$$

where ρ is the mean density of the gas, and F is a complex function of r_0/δ that we will show later in this section. Using Eqs. (2) and (3), the pressure $p(0)$ and the pressure gradient $dp(0)/dx$ at the middle ($x=0$) are, respectively, written as

$$p(0) = p_+ + p_- \tag{4}$$

and

$$v(0) = \frac{iF}{\omega\rho}\left.\frac{dp}{dx}\right|_{x=0} = \frac{kF}{\omega\rho}(p_+ - p_-). \tag{5}$$

In order to relate $p(0)$ and $v(0)$ to the pressures $p_A = p(-\Delta x/2)$ and $p_B = p(\Delta x/2)$ measured by two pressure sensors A and B, we insert $x = \pm\Delta x/2$ into Eq. (2) and, obtain the sum $p_+ + p_-$ and the difference $p_+ - p_-$ as

$$p_+ + p_- = \frac{p_A + p_B}{2\cos(k\Delta x/2)} \tag{6}$$

and

$$p_+ - p_- = \frac{(p_A - p_B)}{2i\sin(k\Delta x/2)}, \tag{7}$$

respectively.

After combining Eq. (4) with Eq. (6), and also Eq. (5) with Eq. (7), the pressure $p(0)$ and the velocity $v(0)$ are expressed using measured pressures p_A and p_B. Finally, the acoustic intensity $I(0)$ is given through Eq. (1) by

$$I = \frac{1}{8\omega\rho}\{\mathrm{Im}[H](|p_A|^2 - |p_B|^2) + 2\,\mathrm{Re}[H]|p_A||p_B|\sin\theta\}, \tag{8}$$

using

$$H = \frac{kF}{\cos(\tilde{k}\Delta x/2)\sin(k\Delta x/2)}, \tag{9}$$

where Im[ ] represents the imaginary part, and $\theta = \arg[p_A/p_B]$ represents the phase lead of p_A relative to p_B. In this way, the acoustic intensity I is determined through measurements of pressure amplitudes $|p_A|$ and $|p_B|$, and their phase difference θ, without involving velocity measurements.

Different from the two-sensor method on the basis of the boundary layer approximation,[10] we use the exact solutions k and F (Ref. 18) for small-amplitude waves in a cylindrical duct given by

$$k = -ik_0\sqrt{\frac{J_0(i^{3/2}\sqrt{2}r_0/\delta)}{J_2(i^{3/2}\sqrt{2}r_0/\delta)}}\sqrt{\gamma + (\gamma-1)\frac{J_2(i^{3/2}\sqrt{2\sigma}r_0/\delta)}{J_0(i^{3/2}\sqrt{2\sigma}r_0/\delta)}} \tag{10}$$

and

$$F = 1 - \frac{2J_1(i^{3/2}\sqrt{2}r_0/\delta)}{i^{3/2}(\sqrt{2}r_0/\delta)J_0(i^{3/2}\sqrt{2}r_0/\delta)}, \tag{11}$$

where γ and σ are the specific heat ratio and Prandtl number, J_n is the nth order complex Bessel function, and k_0 is the wavenumber in free space given by ω divided by the adiabatic speed of sound. Equations (10) and (11) are valid for small-amplitude sound waves in gases contained in a long rigid tube with constant circular cross section. The original two-sensor method that Fusco *et al.* developed uses approximate solutions k' and F' given by

$$k' = k_0\left\{1 + \frac{1-i}{2}\frac{\delta}{r_0}\left(1 + \frac{\gamma-1}{\sqrt{\sigma}}\right)\right\} \tag{12}$$

and

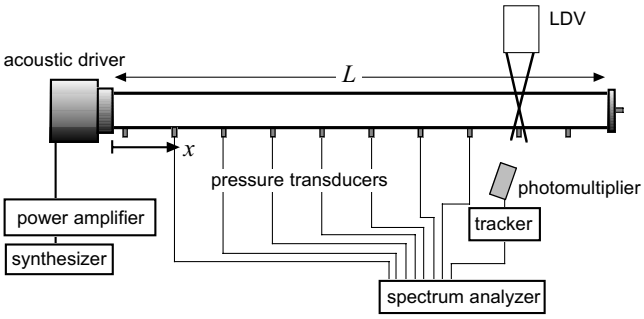


FIG. 2. Schematic illustration of the present experimental setup. The axial coordinate x is directed from the acoustic driver to the closed end.

$$F' = 1 - \frac{1-i}{r_0/\delta}, \tag{13}$$

respectively. Both k and F asymptotically approach k' and F' in the limit of $r_0/\delta \gg 1$, respectively.

Exact solutions k and F have much more complicated expression than k' and F', but there is no practical difficulties in the evaluation of them. We can easily handle the Bessel functions with complex arguments with commercially available software such as MATHEMATICA and MATLAB.

We experimentally test the applicability of the original two-sensor method developed by Fusco *et al.* for wide ducts (r_0/δ=3.76–82). Also, we show, using the narrow duct having r_0/δ=1.3, that the modified two-sensor method that uses k and F extends the applicability of the two-sensor method to narrow duct range.

III. EXPERIMENTAL PROCEDURE

A. Experimental setup

Figure 2 shows a schematic of the present experimental setup. A Pyrex glass cylindrical resonator, closed by a rigid plate at one end and driven at the other end, was filled with atmospheric air at room temperature (284 K). Short copper ducts with a length of 10 mm and an inner radius of 1 mm were attached to the resonator wall, through which the pressure was measured. The number of ducts and the spacing between them were varied by the resonator used. The origin of an axial coordinate x was placed at the driver end of the resonator. Monofrequency sinusoidal voltages were fed from a synthesizer to the acoustic driver (Electro-Voice, ID60DT) through a power amplifier (Yamaha, P1000S). The resonator radius r_0 and the driving frequency f ($=\omega/2\pi$) were chosen as listed in Table I. Thus, five different values of r_0/δ were achieved covering from wide to narrow duct ranges.

TABLE I. Experimental conditions, where L denotes the resonator duct length.

r_0/δ	r_0 (mm)	f (Hz)	L (m)	
58.6	10.5	148.5	1.0	Fundamental
82.8	10.5	298.5	1.0	Second mode
11.9	2.0	168.0	4.3	
3.76	2.0	16.8	4.3	
1.3	2.0	2.0	4.3	

B. Intensity measurements

1. Two-sensor method

Acoustic pressure $p(x)$ was measured by a series of small pressure transducers (JTEKT, DD102-1F) flush mounted on the resonator wall. The natural frequency f_0 of the present transducer is about 3 kHz, and our experimental frequency was always below $f_0/10$. Linear sensitivity was used for the pressure transducers. A multichannel spectrum analyzer (ONOSOKKI, DS-2000) was used to determine the amplitude and the phase of pressures at multiple locations. Prior to the experiments, all the transducers were simultaneously mounted on the rigid end of a dummy resonator, and the sensitivity was calibrated with each other at the frequencies in Table I.

In the two-sensor method, a pair of pressures is needed to determine I at the middle of them. Therefore, using the multiple numbers of pressure transducers makes it possible to determine I at different positions, and so, the axial distribution of I can be quite easily obtained in a single measurement of pressures.

2. Direct method

Axial acoustic particle velocity $U(x)$ on the central axis of the resonator was measured using a LDV. Two laser beams emitted from the identical laser source were crossed inside the glass duct filled with air at ambient pressure and the appropriate amount of cigarette smoke as seeding particles. The intensity of the scattered light was detected by a photomultiplier and was sent to a tracker-type processor. The processor converts the frequency of the intensity variation to the voltage that is proportional to the instantaneous velocity. The voltage signal from the processor was monitored using the spectrum analyzer together with the pressure measured at the same x. The time delay of 2.7×10^{-5} s caused by the processor as well as that associated with the pressure measurements were taken into account when the phase of the velocity was evaluated.[19]

The measured velocity $U(x)$ was confirmed to be in a laminar-flow regime, so the amplitude and the phase of the radial average velocity $V(x)$ were determined from $U(x)$,[9] using the theoretical result of the laminar flow theory instead of measuring the radial profile. The acoustic intensity is determined from Eq. (1) as

$$I = \tfrac{1}{2}|p||v|\cos\phi, \tag{14}$$

where ϕ denotes the phase lead of v to p. The acoustic intensity I determined by the direct method is used as a reference for the intensity I determined by the two-sensor method.

IV. RESULTS AND DISCUSSION

A. Wide duct range

Quality factor Q of a resonator filled with air is approximately given by $Q \simeq r_0/(1.5\delta)$.[9,20] An infinitely high Q value results in a pure standing wave field in the resonator, but as the Q decreases, a fraction of the traveling wave component increases in the acoustic field. Figures 3(a) and 3(b) show the

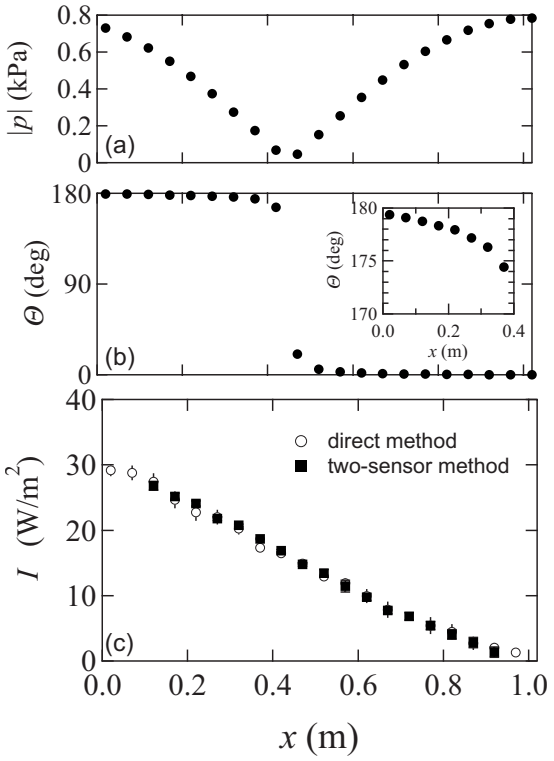


FIG. 3. Acoustic field in the resonator with $r_0/\delta=58.6$. Axial profile of (a) $|p|$ and (b) Θ, and (c) the acoustic intensity I. The open circles (○) in (c) represent I determined by the direct method, whereas the solid squares (■) represent I by the two-sensor method.

axial profile of the pressure amplitude $|p|$ and the phase $\Theta(x)=\arg[p(x)/p_e]$ in a resonator ($r_0=10.5$ mm) driven at the fundamental resonance frequency ($f=148.5$ Hz), where p_e represents the acoustic pressure at the closed end. The parameter $r_0/\delta=58.6$ results in $Q=39$ of this resonator, which is sufficiently high to produce a standing-wave-like acoustic field. Hence, the axial distribution of $|p|$ is close to a rectified cosine wave, and the slope of $\Theta(x)$ is very small except the vicinity of the pressure node ($x\simeq 0.45$ m). A small but finite fraction of the traveling wave component is seen in a nonzero slope of $\Theta(x)$ in the inset of Fig. 3(b).

Figure 3(c) shows the acoustic intensity I obtained by the two-sensor method when the pairs of pressures separated by the distance $\Delta x=0.2$ m are used. We used k' and F' in Eqs. (12) and (13) for the derivation of I in this resonator, according to the ordinary prescription of the two-sensor method. It can be quickly tested whether the experimental I is plausible, by confirming that I monotonically decreases to zero at the closed end, as shown in Fig. 3(c). This is because the velocity and I should become zero at the rigid end, and also because the slope of I should be negative due to the viscous and thermal attenuations. The acoustic intensity I obtained by the direct method was also plotted in Fig. 3(c) as a reference. An excellent agreement was obtained between them. The acoustic intensity I obtained with $\Delta x=0.1$, 0.3, and 0.4 m was also found to agree with I determined by the direct method. Thus, it is clear that the two-sensor method is applicable to the wide-tube resonator with $r_0/\delta=58.6$.

In order to achieve lower r_0/δ values, the duct with $r_0=2.0$ mm was tested. The gas column was driven at f

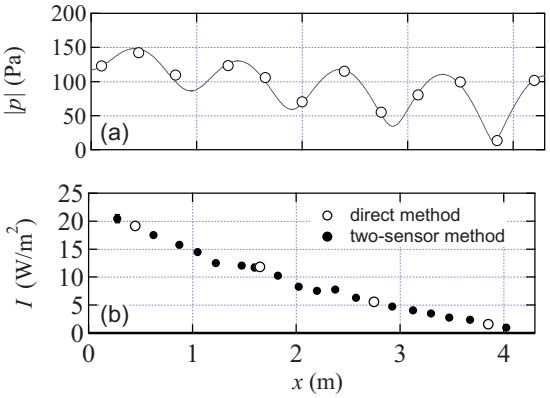


FIG. 4. (Color online) Axial profile of (a) $|p|$ and (b) I when $r_0/\delta=11.9$. The open and solid circles represent I by the direct method and that by the two-sensor method, respectively. The solid line in (a) is a guide to the eye. Error of I is within the size of their symbols.

$=168$ Hz, resulting in $r_0/\delta=11.9$. A lower Q value (=7.9) of this duct increases the ratio of the traveling wave to the standing wave components. As a result, $|p|$ shown in Fig. 4(a) is significantly different from that in a pure standing wave field. The pairs of pressures with $\Delta x=0.35$ and 1.5 m, together with k' and F', were used for the estimation of v and I. We found a good agreement between v measured with the LDV and that determined by the pressures. As a result, the intensities by the two methods agreed well with each other, as shown in Fig. 4(b). This fact assures that the two-sensor method proposed by Fusco *et al.* is applicable also when $r_0/\delta=11.9$.

We further decreased r_0/δ down to 3.76 by decreasing f by a factor of 10. The results obtained by the direct method and the two-sensor method ($\Delta x=0.7$ m) are plotted in Fig. 5. The acoustic intensities obtained by these two methods agree with each other, with the relative error of 5% at $x=0.45$. This fact means that Fusco's method based on the boundary layer approximation is practically applicable even when $r_0/\delta=3.76$. So, this method can cover a quite wide range of r_0/δ. Before showing the test results with a further lower r_0/δ in the next section, it would be beneficial to note pitfalls that we encountered when choosing the distance of the two pressure sensors.

The second term of I in Eq. (8) has a factor of $\sin\theta$, so the relative error of this term is very sensitive to the error of θ when θ is close to $\pm n\pi$, where n is an integer. Hence, it is important to locate the sensors so that a sufficient phase difference is achieved between them. In high Q resonators, the axial distribution of $\Theta(x)$ becomes like a step function, as

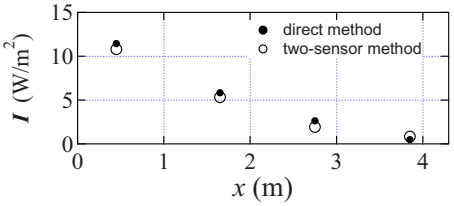


FIG. 5. (Color online) Axial profile of the acoustic intensity I when $r_0/\delta=3.76$. The solid (●) and open (○) circles represent I obtained by the direct method and by the two-sensor method, respectively.

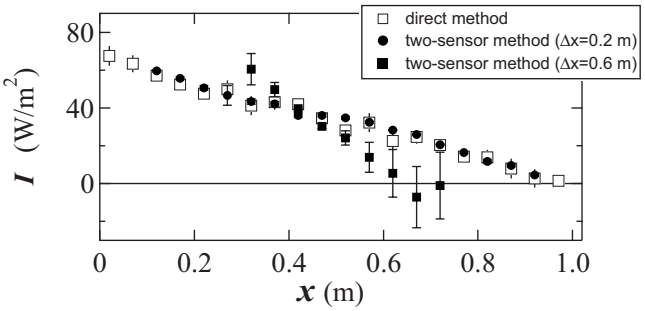


FIG. 6. Axial profile of the acoustic intensity I in the resonator with $r_0/\delta=82.8$. The open squares (□) represent I determined by the direct method, whereas the solid circles (●) and squares (■) represent I determined by the two-sensor method with the use of pressures separated by $\Delta x=0.2$ and 0.6 m, respectively.

shown in Fig. 3(b). This means that the distance Δx between the sensors can become a considerable fraction of the resonator. In our experiments, we only used the pair of pressures that satisfies $|\theta-n\pi|>1.0°$, so as to decrease the error of I within the size of the symbols in the figures.

We should also pay attention to the function H in Eq. (8), when choosing a pair of sensors having appropriate distance Δx.[16,21] Figure 6 shows the acoustic intensity I in the resonator ($r_0=10.5$ mm) driven at the second harmonic frequency ($f=298.5$ Hz). The data plotted by the solid circles (●) were obtained using the pairs of sensors with $\Delta x=0.2$ m, which agree well with I determined by the direct method (□). However, when Δx was increased to 0.6 m, it was found that the deviation of I by the two-sensor method from the true values became significantly large. As explained below, the Δx-dependence of the function H is responsible for the error with $\Delta x=0.6$ m.

Figure 7 shows the real part Re[H] and the imaginary part Im[H] of the function H as a function of the distance Δx. It is shown that Re[H] and Im[H] greatly change with Δx particularly when Δx approaches $n\lambda/2$, where n is a positive integer, and λ is the wavelength. This means that I is very sensitive to slight differences of Δx; large errors are easily induced by small errors of Δx. Similar error of I is unavoidable when errors are present in the constants included in k, such as the adiabatic speed of sound, regardless of whether the boundary layer approximation is valid. So, it is better to avoid $\Delta x=n\lambda/2$ ($n=1,2,3,\ldots$), but to choose Δx near $\lambda/4+(n-1)\lambda/2$, which is best suited to the two-sensor method.

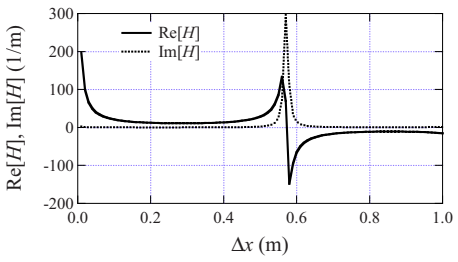


FIG. 7. (Color online) Real part Re[H] (solid line) and imaginary part Im[H] (broken line) as a function of Δx. Here k' and F' are used.

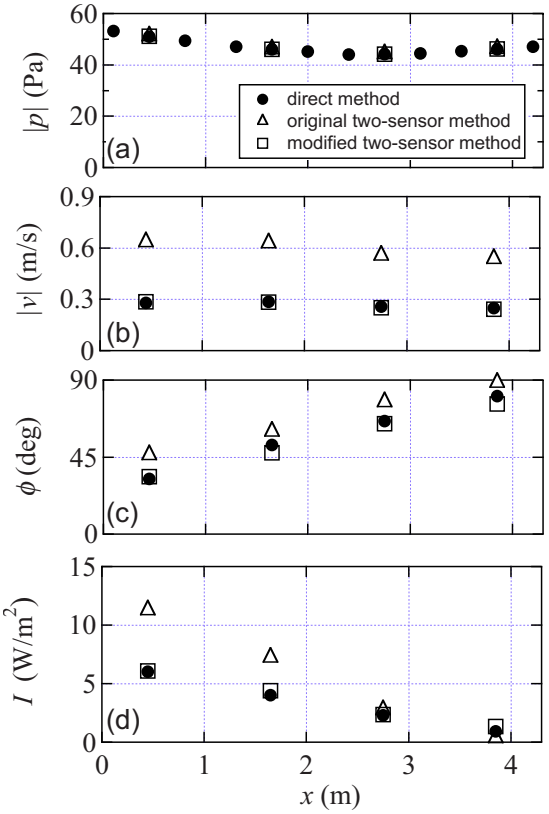


FIG. 8. (Color online) Axial profile of $|p|$, $|v|$, ϕ, and I when $r_0/\delta=1.3$. The open triangles (△) represent the data obtained by the original two-sensor method developed by Fusco *et al.*, and the open squares (□) represent those by the modified two-sensor method using k and F. The solid circles (●) represent the data obtained by the direct method.

B. Narrow duct range

To test the two-sensor method when $r_0/\delta=r_0\sqrt{\omega/2\nu}$ is close to 1, the same glass duct with $r_0=2.0$ mm was used as in the preceding section, but the frequency was further reduced to $f=2.0$ Hz. As a result, the ratio $r_0/\delta=1.3$ was achieved. With the duct closed by a rigid plate, the phase $\Theta(4.2)$ at the driver end was found to be 2.6°, meaning that a very large Δx between the sensors was needed to satisfy $|\theta|>1.0°$. So, in this particular experiment, we replaced the rigid plate with a rubber balloon, to increase the traveling wave component.[17]

We determined p and v using the pressures separated by the distance $\Delta x=0.7$ m. We adopted the original two-sensor method, namely, using k' and F' obtained under the boundary layer approximation. The estimated $|p|$, $|v|$, and their phase difference ϕ are plotted by open triangles (△) in Figs. 8(a)–8(c), respectively. For comparison, those determined by the direct measurements are also plotted by solid circles (●). It is shown that the pressure amplitude $|p|$ is in good agreement with the true value, but v estimated by the two-sensor method is apparently different from that directly measured by the LDV; $|v|$ is 2.2 times larger than the true value, and the phase ϕ leads the true v by 12°. Since I depends on both $|v|$ and ϕ through the factor $|v|\cos\phi$ in Eq. (14), the two-sensor method gives significantly different I from that by the direct method, which is shown in Fig. 8(d). Thus, we conclude that the two-sensor method using k' and F' becomes inappropriate when r_0/δ decreases to 1.3.

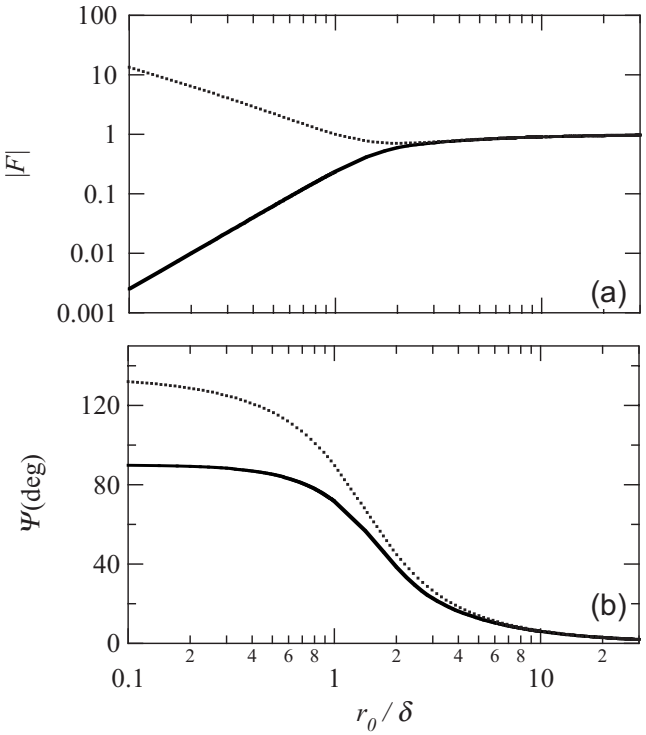


FIG. 9. The complex factor $F=|F|e^{i\Psi}$ as a function of the ratio r_0/δ. The dotted curves represent F' obtained under the boundary layer approximation.

In order to obtain the correct I by the two-sensor method, we modified the original method by replacing k' and F' with k and F that are given in Eqs. (10) and (11). The results determined by the modified two-sensor method are shown by open squares (□) in Figs. 8(a)–8(d). We see that $|v|$, ϕ, and I obtained by the modified two-sensor method agree with those by the direct method. This result shows that the present formulation using k and F makes it possible to successfully measure I in narrow ducts.

The complex wavenumber k is theoretically derived by Tijdeman,[18] and is experimentally verified including the narrow duct range from $r_0/\delta=10^{-2}$ to 10 recently.[22] As is shown in these literatures, k' obtained under the boundary layer approximation deviates from k below $r_0/\delta\sim 4$. However, in the present experiment, the use of k' does not cause serious difference. As shown in Fig. 8(a), the amplitude $|p|$ estimated by the two-sensor methods (△ and □) falls onto the directly measured data, irrespective of whether k' is substituted for k. Such agreement is attributable to the fact that Δx is much smaller than the wavelength in this experiment; if Δx is comparable to the wavelength, the use of k instead of k' would become necessary.

Large discrepancy of I in the original two-sensor method comes from the complex factor F'. We plotted F and F' in Figs. 9(a) and 9(b) as a function of r_0/δ. The difference between the absolute values of F and F' rapidly grows below $r_0/\delta\sim 4$, and the difference between their arguments becomes a few degrees with $r_0/\delta<10$. When $r_0/\delta=1.3$, the ratio of their absolute values reaches 2.2 and the phase difference becomes 12°, respectively. These differences result in the large deviation of v having the factor F in Eq. (5). It should be noted that the original two-sensor method fortuitously worked well when $r_0/\delta=3.76$. The difference of the argument reaches 2.5° between F and F' in this case, whereas $|F'|=1.02|F|$. The deviation of the phase ϕ can be critical in the measurement of $I(\propto\cos\phi)$, when ϕ is close to the standing wave phase $(\pm\pi/2)$.[9] Indeed, in Fig. 5, the relative error of I exceeds 60% at $x=3.85$ m where ϕ reaches 85°, while it remains within 5% at $x=0.45$ m where $\phi=28°$. Thus, it is strongly recommended to use k and F to obtain the best result in the measurement of I in ducts.

V. SUMMARY

In this paper, the acoustic intensity I in cylindrical ducts was determined by the two-sensor method and also by the direct method measuring pressure and velocity. From the comparison with the acoustic intensity obtained by the direct method, the applicability of the two-sensor method was tested with r_0/δ values from 1.3 to 82.8. The two-sensor method is useful for wide ducts, but it was found to become inappropriate for a narrow duct with $r_0/\delta=1.3$. We developed the two-sensor method to include a narrow duct range by modifying the formulation of the method. It is shown that the two-sensor method that we developed enables quick and accurate evaluation of the acoustic intensity I in lossy ducts from narrow to wide duct ranges.

[1] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Sewickley, PA, 2002).
[2] A. Tominaga, "Thermodynamic aspects of thermoacoustic theory," Cryogenics **35**, 427–440 (1995); *Fundamental Thermoacoustics* (Uchida Roukakuho, Tokyo, Japan 1998).
[3] T. Yazaki and A. Tominaga, "Measurement of sound generation in thermoacoustic oscillations," Proc. R. Soc. London, Ser. A **454**, 2113–2122 (1998).
[4] T. Yazaki, A. Iwata, T. Maekawa, and A. Tominaga, "Traveling wave thermoacoustic engine in a looped tube," Phys. Rev. Lett. **81**, 3128–3131 (1998).
[5] T. Yazaki, T. Biwa, and A. Tominaga, "A pistonless Stirling cooler," Appl. Phys. Lett. **80**, 157–159 (2002).
[6] Y. Ueda, T. Biwa, U. Mizutani, and T. Yazaki, "Acoustic field in a thermoacoustic Stirling engine having a looped tube and resonator," Appl. Phys. Lett. **81**, 5252–5254 (2002); "Experimental studies of a thermoacoustic Stiriling prime mover and its application to a cooler," J. Acoust. Soc. Am. **115**, 1134–1141 (2004).
[7] T. Biwa, Y. Tashiro, U. Mizutani, M. Kozuka, and T. Yazaki, "Experimental demonstration of thermoacoustic energy conversion in a resonator," Phys. Rev. E **69**, 066304 (2004).
[8] H. Bailliet, P. Lotton, M. Bruneau, V. Gusev, J. C. Valiere, and B. Gazengel, "Acoustic power flow measurement in a thermoacoustic resonator by means of laser Doppler anemometry (L. D. A.) and microphonic measurement," Appl. Acoust. **60**, 1–11 (2000).
[9] T. Biwa, Y. Ueda, H. Nomura, U. Mizutani, and T. Yazaki, "Measurement of the Q value of an acoustic resonator," Phys. Rev. E **72**, 026601 (2005).
[10] A. M. Fusco, W. C. Ward, and G. W. Swift, "Two-sensor power measurements in lossy ducts," J. Acoust. Soc. Am. **91**, 2229–2235 (1992).
[11] G. W. Swift, D. L. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[12] G. Petculescu and L. A. Wilen, "Traveling-wave amplification in a variable standing wave ratio device," ARLO **3**, 71–76 (2002).
[13] D. L. Gardner and G. W. Swift, "A cascade thermoacoustic engine," J. Acoust. Soc. Am. **114**, 1905–1919 (2003).
[14] J. Y. Chung and D. A. Blaster, "Transfer function method of measuring in-duct acoustic properties. I. theory," J. Acoust. Soc. Am. **68**, 907–913 (1980); "Transfer function method of measuring in-duct-acsoutic properties. II. experiment," J. Acoust. Soc. Am. **68**, 913–921 (1980).
[15] W. T. Chu, "Extension of the two-microphone transfer function method for

impedance tube measurements," J. Acoust. Soc. Am. **80**, 347–348 (1986).
[16]A. F. Seybert and D. F. Ross, "Experimental determination of acoustic properties using a two-microphone random-excitation technique," J. Acoust. Soc. Am. **61**, 1362–1370 (1977).
[17]T. Biwa, Y. Tashiro, H. Nomura, Y. Ueda, and T. Yazaki, "Acoustic intensity measurement in a narrow duct by a two-sensor method," Rev. Sci. Instrum. **78**, 086110 (2007).
[18]H. Tijdeman, "On the propagation of sound waves in cylindrical tubes," J. Sound Vib. **39**, 1–33 (1975).
[19]T. Biwa, Y. Ueda, T. Yazaki, and U. Mizutani, "Work flow measurements in a thermoacoustic engine," Cryogenics **41**, 305–310 (2001).
[20]G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[21]H. Bodén and M. Åbom, "Influence of errors on the two-michrophone method for measuring acoustic properties in ducts," J. Acoust. Soc. Am. **79**, 541–549 (1986).
[22]T. Yazaki, Y. Tashiro, and T. Biwa, "Measurement of sound propagation in narrow tubes," Proc. R. Soc. London, Ser. A **463**, 2855–2862 (2007).

Reconstruction of material properties profiles in one-dimensional macroscopically inhomogeneous rigid frame porous media in the frequency domain

L. De Ryck[a] and W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, B-3001 Heverlee, Belgium

P. Leclaire
Laboratoire de Recherche en Mécanique et Acoustique, ISAT/Université de Bourgogne, 49 rue Mademoiselle Bourgeois, B.P.31-58027 Nevers cedex, France

J. P. Groby
Département de Recherche en Electromagnétisme, Laboratoire des Signaux et Systèmes, UMR8506 CNRS/Supelec/UPS11, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

A. Wirgin
Laboratoire de Mécanique et d'Acoustique, UPR 7051 CNRS, 31 chemin Joseph Aiguier, 13009 Marseille, France

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

(Received 4 April 2008; revised 16 June 2008; accepted 17 June 2008)

The present paper deals with the inverse scattering problem involving macroscopically inhomogeneous rigid frame porous media. It consists of the recovery, from acoustic measurements, of the profiles of spatially varying material parameters by means of an optimization approach. The resolution is based on the modeling of acoustic wave propagation in macroscopically inhomogeneous rigid frame porous materials, which was recently derived from the generalized Biot's theory. In practice, the inverse problem is solved by minimizing an objective function defined in the least-square sense by the comparison of the calculated reflection (and transmission) coefficient(s) with the measured or synthetic one(s), affected or not by additive Gaussian noise. From an initial guess, the profiles of the x-dependent material parameters are reconstructed iteratively with the help of a standard conjugate gradient method. The convergence rate of the latter and the accuracy of the reconstructions are improved by the availability of an analytical gradient. [DOI: 10.1121/1.2959734]

PACS number(s): 43.60.Pt, 43.20.Fn, 43.40.Sk [KA] Pages: 1591–1606

I. INTRODUCTION

Porous materials are encountered in a large domain of application in building acoustics (concrete walls, rockwools, plastic foams), in aeronautic, in geophysics, in petroleum prospection, in medicine (osteoporosis diagnose)… The resolution of the inverse scattering problem for the characterization of the parameters of rigid frame porous materials is of large importance, particularly when they are spatially varying. The parameters of interest in rigid frame porous materials are the porosity ϕ, which is the ratio of the fluid volume to the total sample volume; the tortuosity τ_∞, which is a parameter describing the change in magnitude and direction of the fluid microvelocity due to the curliness of the pores; the characteristic viscous and thermal lengths Λ and Λ', which are illustrating the geometry of the pores through the viscous and thermal losses; and the flow resistivity R_f, which is the ratio of the fluid viscosity η to the fluid permeability κ_f.

Wave propagation in homogeneous porous materials was mainly studied after Biot's theory[1–4] and latter contributions.[5–8] However, the equations of motion for macroscopically inhomogeneous porous materials were only recently derived from the alternative formulation of Biot's theory.[3,4,9,10] In particular, when the assumption of rigid frame is valid,[5–7,11] i.e., when the saturating fluid is light, such as air, and the frame is not moving, the equations of motion reduce to those of an equivalent inhomogeneous fluid with spatial and frequency dependent effective density and bulk modulus.[9,10,12] The frequency band suitable to this approximation is bounded at high frequency by the diffusion limit (when the wavelength is of the order of, or smaller than the pore size), and at low frequency by the Biot characteristic frequency, below the which the skeleton may vibrate. Under these conditions, the reflected and transmitted pressure fields

[a]Author to whom correspondence should be addressed. Electronic mail: laurent.deryck@fys.kuleuven.be

as well as the internal pressure field can be numerically determined[9,10] by means of the wave splitting and transmission Green's functions method.[13] This method is employed in the present paper to generate synthetic data to feed the optimizer solving the inverse scattering problem. For multilayered materials, the analytic calculation of the scattered fields is preferred.

Several methods have been developed to characterize homogeneous porous samples, or layered homogeneous porous samples. One is based on the measurement of the change in the fluid volume of a chamber without or with the sample (porosimeter for the determination of the porosity)[14]; another one concerns the simultaneous measurement of the pressure drop accross a sample and the rate of fluid flow through a porous sample (flowmeter for the determination of the flow resitivity)[15]; more recent methods deal with ultrasonic measurements in both time and frequency domains to retrieve the following parameters: the tortuosity,[16] the viscous[17] and thermal characteristic lengths (the $Q\delta$ method),[11] the porosity and the tortuosity simultaneously,[18,19] and finally the porosity, the tortuosity, and the viscous characteristic length simultaneously.[20] The complexity of the problem is hardened by the different sensitivity of each parameter on the physical quantities that are measured and to the frequency of the solicitation. For example, both density and bulk modulus do not depend on R_f in the asymptotic high frequency model.

The aim of the following material is to reconstruct the depth profiles of all or some material parameters from the measurement of the acoustic field scattered (more specifically: reflected and/or transmitted) by a sample of known thickness. To solve such inverse problems, the Wave Splitting method was extensively studied in the time domain.[21–26] More recently, similar methods, together with an optimization approach, were developed in the frequency domain.[27,28] The latter are more suitable to model wave propagation in rigid frame macroscopically inhomogeneous porous materials in all its complexity, i.e., over a large frequency range for which the five parameters have to be accounted for. Moreover, this allows us to work with band-limited data, which is more appropriate when one has to deal with measured data. The inverse scattering problem considered here is unidimensional (depth profiling) and is solved iteratively via a minimization approach based on a classical conjugate gradient method (CG). This method minimizes an objective function which compares measured (or synthetic) reflection coefficients with numerically estimated ones obtained by the frequency domain wave splitting and invariant imbedding technique.[24,27] The availability of the analytical gradient of the objective function allows to accelerate the rate of convergence of the optimization method. The choice of the invariant imbedding technique instead of the transmission Green's function method is motivated by the use of the reflection coefficient and optionally the transmission coefficient, but also by the necessity to avoid computational traps and errors due to the so-called "back marching effect," also known as the "inverse crime."[29] The latter arises when the synthetic data are generated with a direct solver identical or very similar to the one employed in the minimization method. When the invariant imbedding (inverse problem) and transmission Green's functions technique (direct problem) are used together with the wave splitting method, Gaussian noise is added to the data to avoid the inverse crime.

The last part of this paper deals with profile reconstructions of one or several material parameters. The accuracy and computation time of the reconstructions are discussed.

II. MACROSCOPICALLY INHOMOGENEOUS RIGID FRAME POROUS MATERIALS

In this section, the acoustic wave propagation model in macroscopically inhomogeneous rigid frame porous materials is presented.[9,10,12] The theory of wave propagation in homogeneous porous materials was initially given by Biot.[1–4] In most of the plastic foams saturated by a light fluid like air, the rigid frame assumption is valid so that an acoustic excitation impinging on a porous sample induces wave propagation only in the fluid phase. Therefore the viscothermal effects taking place in the pore channels are included in an effective density ρ_e and an effective bulk modulus K_e of a so-called equivalent fluid.[7]

The model of rigid frame has been recently extended to macroscopically inhomogeneous porous media.[9,10,12] In the frequency domain, the Helmholtz equation in terms of the fluid pressure p inside the equivalent inhomogeneous fluid is[9,10]

$$\nabla \cdot \left(\frac{1}{\rho_e(x,\omega)} \nabla p \right) + \frac{\omega^2}{K_e(x,\omega)} p = 0, \tag{1}$$

where ω is the angular frequency and ∇ the nabla operator. Attenuation, viscothermal losses and dispersion are accounted for in the effective density and bulk modulus. Effective sound speed and characteristic impedance are $c_e(x,\omega) = \sqrt{K_e(x,\omega)/\rho_e(x,\omega)}$ and $Z_e = \rho_e(x,\omega) c_e(x,\omega)$.

In all of this paper, inhomogeneities occur along the x direction and so a unidimensional approach is considered. According to the Biot–Johnson–Allard model[1–7] extended to macroscopically inhomogeneous porous materials:[9,10]

$$\rho_e(x,\omega) = \rho_f \frac{\tau_\infty(x)}{\phi(x)} \left[1 - j \frac{R_f(x)\phi(x)}{\rho_f \tau_\infty(x) \omega} F(x,\omega) \right], \tag{2}$$

$$K_e(x,\omega) = \frac{\gamma P_0 / \phi(x)}{\gamma - (\gamma - 1) \left[1 - j \dfrac{R_f(x)\phi(x)}{\rho_f \tau_\infty(x) \mathrm{P}_r \omega} G(x, \mathrm{P}_r \omega) \right]^{-1}}, \tag{3}$$

wherein γ is the specific heat ratio, ρ_f the saturating fluid density, P_0 the atmospheric pressure and P_r the Prandtl number. The well-defined correction functions of this model are[5–7]

$$F(x,\omega) = \sqrt{1 + j4 \frac{\eta \rho_f \tau_\infty^2(x)}{R_f(x)^2 \phi(x)^2 \Lambda^2(x)} \omega}, \tag{4}$$

$$G(x, \mathrm{P}_r \omega) = \sqrt{1 + j4 \frac{\eta \rho_f \tau_\infty^2(x)}{R_f(x)^2 \phi(x)^2 \Lambda'^2(x)} \mathrm{P}_r \omega}. \tag{5}$$

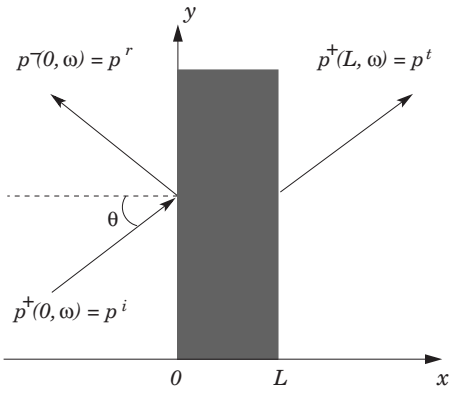


FIG. 1. Slab of porous material of thickness L. p^i is the incident signal, p^r is the reflected signal, and p^t is the transmitted signal. The angle of incidence is denoted by θ. In a homogeneous medium, p^- are the backward propagating waves and p^+ are the forward propagating waves.

The five so-called acoustical parameters $\phi(x)$, $\tau_\infty(x)$, $\Lambda(x)$, $\Lambda'(x)$, and $R_f(x)$ described in the introduction become here space-dependent functions.[9,10] Without addition of shape factors, the thermal and viscous permeabilities are assumed to be equal so that the flow resistivity R_f appears in both correction functions.[6,7]

The wave splitting and invariant imbedding method chosen to solve both the direct and inverse scattering problem is solely based on the equations of motion (6):[9,10]

$$\frac{d}{dx}\begin{pmatrix} p \\ \phi(x)V \end{pmatrix} = \begin{pmatrix} 0 & -j\omega\rho_e(x,\omega) \\ -\dfrac{j\omega}{K_e(x,\omega)} & 0 \end{pmatrix}\begin{pmatrix} p \\ \phi(x)V \end{pmatrix}, \tag{6}$$

where V is the fluid particle velocity. This set of equations is equivalent to the Helmholtz equation (1).

III. STATEMENTS OF THE PROBLEM AND DIRECT SCATTERING SOLUTION

In all the following paper, a macroscopically one-dimensional (1D) inhomogeneous porous slab of thickness L, solicited by an incident plane wave propagating initially in the air (the saturating fluid here), is considered. This is depicted in Fig. 1.

The wave splitting method was mainly used to solve both direct and inverse scattering problems in several fields of physics, especially in electromagnetism, initially in the time domain[21–23,30] and then in the frequency domain.[9,10,13,27] In order to avoid convolution products in time and fractional derivatives,[18,20] the direct and inverse scattering problems are treated in the frequency domain. Also, the Fourier transform $p(x,\omega)$ of the measurable pressure field $p(x,t)$ is used. The inverse Fourier transform is chosen such that

$$p(x,t) = \int_{-\infty}^{\infty} p(x,\omega)\exp(j\omega t)d\omega. \tag{7}$$

The inverse scattering problem consists in the retrieval of the profile of the five acoustical parameters from the measurement of the pressure waves reflected by an inhomogeneous porous sample. The solution of the direct scattering problem is used to improve each profile at each step of the iterative optimization approach. In order to be able to simultaneously reconstruct more than one depth-dependent material property, enough information must be provided. Several frequencies are used in order (I) to regularize the inverse scattering problem, (II) to use a large enough frequency band for most of the acoustic parameters to have an influence on the scattered field, and (III) to be able to recover from the "low frequency" content of the scattered field the average values of the acoustic parameters and from its "high frequency" content their spatial variations, which can be sharp in case of layered medium. On the other hand, the problem is much more complicated than the one usually treated in homogeneous case because, for instance, the porosity and tortuosity are space dependent functions and cannot be "simply" recovered from the instantaneous reflection at the first interface of the sample. The additional information available from the wave front during its propagation inside the whole sample has to be caught. To increase the amount of information, several operations are applicable: (I) For the simultaneous reconstruction of two profiles, a two-sided normal reflection measurement can be used, i.e., the recording of both the reflected data at $x=0$ after an excitation from a source at $x<0$ and the reflected data at $x=L$ after an excitation in the opposite direction from $x>L$; (II) For the simultaneous reconstruction of one to four depth profiles, several two-sided reflection measurements at normal incidence can be done by imposing or not a rigid wall condition on the opposite side of the solicitation; (III) Another classical way to solve the problem of reconstructing many depth profiles is to use acoustic waves impinging on the porous sample at various oblique incidence. The last operation is followed and as many profiles as angles of incidence should normally be reconstructed.

The direct scattering problem consists of determining the scattered, i.e., reflected and transmitted, pressure field due to the presence of a macroscopically inhomogeneous porous slab. Therefore, the aim is to solve the system of equations (6) for an incident acoustic excitation from a position $x<0$. The solution is obtained through the wave splitting and invariant imbedding approach[9,10,12,27], which requires boundary conditions. They are found via the characteristic impedances of both surrounding media for $x<0$ and $x>L$. Here, for all $x<0$, the medium is a homogeneous free fluid (air). Its characteristic impedance is denoted by $Z_0=\rho_f c_f$ with c_f the sound speed in the free fluid. For all $x>L$, the medium can be either air or a rigid wall. Its impedance is denoted by Z_L, which is respectively Z_0 in the first case and the infinity in the second case.

A. Wave splitting transformation

Wave splitting is a method based on the mathematical properties of the wave equation. It had been demonstrated since d'Alembert[31] that the solution of the wave equation in terms of the pressure field $p(x,t)$, which is the inverse Fourier transform of the Helmholtz Eq. (1), can be exactly decomposed into two components $p^+(x,t)$ and $p^-(x,t)$ corresponding to forward and backward propagating waves in a homogeneous medium.[13,21–24,32] The transformation is gener-

alized for inhomogeneous media since it can be seen as an algebraic transformation from the base of functions $[p;\phi(x)V]$ to the base of functions $[p^+;p^-]$. Several transformations are available. In the present case, the so-called vacuum wave splitting tranformation[9,10,27,13] is chosen,

$$\begin{pmatrix} p^+ \\ p^- \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & Z_0 \\ 1 & -Z_0 \end{pmatrix}\begin{pmatrix} p \\ \phi(x)V \end{pmatrix}. \tag{8}$$

This choice is possible because the wave splitting is coupled with an invariant imbedding technique. In a general wave splitting transformation, the characteristic material impedance $Z_e(x,\omega)$ should be used.[21–23] However, the invariant imbedding technique, which will be detailed in the next section, allows a transformation that uses the known characteristic impedance $Z_0=\rho_f c_f$ of the saturating fluid instead.[27,28] As a schematic explanation, one considers that at the initial state, and for all $x<L$, there is only the ambient fluid of characteristic impedance Z_0. Then, layer after layer, the inhomogeneous porous material is constructed from $x=L$ to $x=0$. Therefore, in the present formulation, each nth layer of infinitesimal thickness dx added at $x=L-ndx$ is taken into account as a perturbation of the impedance Z_0 at the point $x=L-ndx$. Continuity conditions from layer to layer are implicitely accounted for.

Applying the transformation (8) to the equations of motion (6) in the frequency domain straightforwardly yields[9,10]

$$\frac{d}{dx}\begin{pmatrix} p^+ \\ p^- \end{pmatrix}(x,\omega) = \begin{pmatrix} -A^+ & -A^- \\ A^- & A^+ \end{pmatrix}\begin{pmatrix} p^+ \\ p^- \end{pmatrix}(x,\omega), \tag{9}$$

with

$$A^{\pm}(x,\omega) = \frac{j\omega}{2}\left[\frac{Z_0}{K_e(x,\omega)} \pm \frac{\rho_e(x,\omega)}{Z_0}\right]. \tag{10}$$

For all $x\leq 0$, the pressures waves $p^+(x,\omega)$ and $p^-(x,\omega)$ are respectively the incident and reflected pressure waves. Similarly, for all $x\geq L$, $p^+(x,\omega)$ corresponds to the transmitted pressure wave. Therefore, the splitted pressure waves $p^{\pm}$ are easily connected to the reflection and transmission coefficients $R(\omega)$ and $T(\omega)$. These coefficients contain information on the inhomogeneities inside the sample, but an extension of these data for each depth x is necessary to take accurately into account the material properties variations between $x=0$ and $x=L$. A continuous approach is done in the following by taking the limit $dx\to 0$.

B. Invariant imbedding method

In the invariant imbedding method, an imbedding geometry is considered, i.e., a subslab between x and L of the total material slab of thickness L. For each subslab, it is assumed that the medium filling the domain between 0 and x is the same as the medium chosen for all $x<0$, e.g., air in the treated case. This allows us to define the x-dependent reflection and transmission coefficients $R(x,\omega)$ and $T(x,\omega)$ as follows:

$$p^-(x,\omega) = R(x,\omega)p^+(x,\omega), \tag{11}$$

$$p^+(L,\omega) = T(x,\omega)p^+(x,\omega). \tag{12}$$

From the above relations, one deduces that $R(0,\omega)=R(\omega)$ and $T(0,\omega)=T(\omega)$ are the usual reflection and transmission coefficients for the total porous slab of length L. These coefficients link the measurable reflected and transmitted pressure fields, respectively denoted by $p^r(\omega)=p^-(0,\omega)$ and $p^t(\omega)=p^+(L,\omega)$, to the incident pressure field $p^i(\omega)=p^+(0,\omega)$ illuminating the porous slab at $x=0$, as in Fig. 1.

Combining Eqs. (11) and (12) with (9) and (10) leads to the imbedding equations for $R(x,\omega)$ and $T(x,\omega)$,

$$\frac{d}{dx}R = 2A^+R + A^-(1+R^2), \tag{13}$$

$$\frac{d}{dx}T = (A^+ + A^-R)T. \tag{14}$$

The first equation is a nonlinear Riccati differential equation for R, which can be solved with a classical fourth order Runge Kutta algorithm. The second one is a linear differential equation for T, which can be solved with the help of the method described in Appendix A, once the Riccati equation is solved for R. The resolution of these equations requires boundary conditions to be satisfied. According to the definition of the imbedding geometry, there is no porous subslab between x and L when $x=L$. This means that there is just a simple interface between the two homogeneous media of characteristic impedances Z_0 and Z_L. These impedances are constants so that the boundary conditions at $x=L$ for the imbedding geometry are

$$R(L,\omega) = \frac{Z_L - Z_0}{Z_L + Z_0}, \tag{15}$$

$$T(L,\omega) = \frac{2Z_L}{Z_L + Z_0}. \tag{16}$$

If the homogeneous medium for all $x>L$ is the same free fluid as the one for all $x<0$, then $Z_L=Z_0$ and the boundary conditions reduce to

$$R(L,\omega) = 0, \tag{17}$$

$$T(L,\omega) = 1. \tag{18}$$

If the homogeneous medium for all $x>L$ is a rigid wall, then $Z_L=\infty$. There is no transmitted waves, and only the imbedding Eq. (13) for $R(x,\omega)$ is used. The boundary condition reduces to

$$R(L,\omega) = 1. \tag{19}$$

Equations (11) and (12) together with the boundary conditions (15) and (16) constitute the direct solver.

C. Notes about the oblique incidence

The oblique incidence implies some slight changes in Eqs. (6)–(14). In the latters, for an angle of incidence θ, the effective bulk modulus K_e, the characteristic impedances Z_0 and Z_e must be replaced by[10]

$$\bar{K}_e = K_e\left(1 - \frac{c_e^2}{c_f^2}\sin^2\theta\right)^{-1}, \tag{20}$$

$$\bar{Z}_0 = \frac{Z_0}{\cos\theta}, \tag{21}$$

$$\bar{Z}_e = Z_e\left(1 - \frac{c_e^2}{c_f^2}\sin^2\theta\right)^{-1/2}. \tag{22}$$

These expressions are extracted from the projection of the wave vector onto the x axis.

In the next section, the optimization method that is employed to solved the inverse problem of the recovery of the profiles is presented. The transmission data are not often used[28] and may even trap the optimizer.[27] Therefore, in the following, only the minimization of the reflection data is detailed.

IV. OPTIMIZATION APPROACH OF THE INVERSE SCATTERING PROBLEM

Let us introduce a five-element vector

$$\mathbf{p} = (\phi(x), \tau_\infty(x), \Lambda(x), \Lambda'(x), R_f(x))^T, \tag{23}$$

where the superscript ^{T} denotes the transpose operation. An objective function $J(\mathbf{p})$ is defined such that[27]

$$J(\mathbf{p}) = \sum_{\theta_{\min}}^{\theta_{\max}} \sum_{\omega_{\min}}^{\omega_{\max}} W_R(\omega)|R(0,\omega;\theta) - R_m(\omega;\theta)|^2, \tag{24}$$

where R_m is the measured reflection coefficient, $R(0,\omega;\theta)$ the reflection coefficient calculated at $x=0$ with the help of our direct solver presented in the previous section, and W_R a non-negative frequency-dependent weighting function.

The objective function is defined such that a summation is performed over the angles of incidence between $\theta_{\min}$ and $\theta_{\max}$, and over a frequency band $[\omega_{\min}, \omega_{\max}]$. The larger the number of angles of incidence and the number of discrete frequencies, the more regularized the minimization problem is.

The optimal values of the elements of $\mathbf{p}$ are obtained when the objective function $J(\mathbf{p})$ is minimal. These optimal values correspond to the profiles that are to be reconstructed. Therefore the optimization approach consists of minimizing the function $J(\mathbf{p})$ from an initial guess of the elements of the vector $\mathbf{p}$. The resolution of our problem, involving the minimization of the function $J(\mathbf{p})$, is hopeless if no other information is available. If the gradient of the objective function is known for all x—analytically or numerically—the convergence to the global minimum is much improved and accelerated.[25,27] This is due to the fact that the gradient provides the best directions to follow in order to converge to a minimum. Moreover, if the gradient is available for all x, the profiles of each parameter, along the sample thickness, are optimizable. Then, well-known iterative minimization routines can be applied.

A. Analytical gradient of the objective function

A small perturbation $\delta\mathbf{p}$ is applied to the material property vector $\mathbf{p}$. By definition, the perturbation of the reflection coefficient is $\delta R = \tilde{R} - R$ where $\tilde{R} = R(x,\omega;\mathbf{p}+\delta\mathbf{p})$ is the solution of the perturbed Eq. (13), which becomes:[27]

$$\frac{d}{dx}\delta R - 2(A^+ + A^- R)\delta R = 2R\,\delta A^+ + (1+R^2)\delta A^- \tag{25}$$

with the perturbed condition (15) reducing to

$$\delta R(L,\omega) = 0 \tag{26}$$

because the impedances $Z_0 = \rho_f c_f$ and Z_L are independent of the material properties, which are solely included in the vector $\mathbf{p}$.

The perturbations of the coefficients $A^\pm$ are deduced from Eq. (10) and are expressed, for any oblique angle of incidence, as

$$\delta A^\pm = \frac{j\omega}{2}[Z_0 \delta \bar{K}_e^{-1} \pm Z_0^{-1}\delta\rho_e], \tag{27}$$

wherein the perturbation of ρ_e [Eq. (2)] and of $\bar{K}_e^{-1}$ [Eq. (20)] yield

$$\delta\rho_e = \frac{\partial\rho_e}{\partial\phi}\delta\phi + \frac{\partial\rho_e}{\partial\tau_\infty}\delta\tau_\infty + \frac{\partial\rho_e}{\partial\Lambda}\delta\Lambda + \frac{\partial\rho_e}{\partial R_f}\delta R_f, \tag{28}$$

$$\delta\bar{K}_e^{-1} = \delta K_e^{-1} + \frac{\sin^2\theta}{\rho_e^2 c_f^2}\delta\rho_e \tag{29}$$

with

$$\delta K_e^{-1} = \frac{\partial K_e^{-1}}{\partial\phi}\delta\phi + \frac{\partial K_e^{-1}}{\partial\tau_\infty}\delta\tau_\infty + \frac{\partial K_e^{-1}}{\partial\Lambda'}\delta\Lambda' + \frac{\partial K_e^{-1}}{\partial R_f}\delta R_f. \tag{30}$$

From Eqs. (2) and (3), the partial derivatives in Eqs. (28) and (30) are reduced to functions in terms of $\phi(x)$, $\tau_\infty(x)$, $\Lambda(x)$, $\Lambda'(x)$ and $R_f(x)$.

The increment of the objective function (24) induced by the perturbation $\delta\mathbf{p}$ takes the form[25,27]

$$\delta J(\mathbf{p}) = J(\mathbf{p}+\delta\mathbf{p}) - J(\mathbf{p}) = 2\,\mathrm{Re}\sum_{\theta,\omega} u(0,\omega;\theta)\,\delta R(0,\omega;\theta), \tag{31}$$

where Re represents for the real part and wherein the auxiliary function $u(x,\omega;\theta)$ is defined such that

$$u(0,\omega;\theta) = W_R(\omega)(R(0,\omega;\theta) - R_m(\omega;\theta))^\star, \tag{32}$$

in which the superscript * denotes the complex conjugate.

Using the boundary conditions (26) at $x=L$, the following integration rule can be applied:

$$-u(0,\omega;\theta)\,\delta R(0,\omega;\theta) = \int_0^L \left[\frac{d}{dx}u(x,\omega;\theta)\,\delta R(x,\omega;\theta)\right]_{x=z} dz, \tag{33}$$

the integrand of which, from Eq. (25), is developed as follows:

$$\frac{d}{dx}(u\,\delta R) = \frac{du}{dx}\delta R + u\frac{d\delta R}{dx} = \left(\frac{du}{dx} + 2(A^+ + A^- R)u\right)\delta R + 2Ru\,\delta A^+ + (1+R^2)u\,\delta A^-. \tag{34}$$

$u(x,\omega;\theta)$ is an arbitrary function, except at $x=0$, so that we can choose it in order to eliminate the dependence on the

unknown perturbation δR in Eq. (34). Thus, the following linear first order differential equation must be satisfied:

$$\frac{du}{dx} + 2(A^+ + A^- R)u = 0, \tag{35}$$

with the boundary conditions at $x=0$ given in Eq. (32). This equation can be solved iteratively, as explained in Appendix A.

Combining Eqs. (33) and (31) with the condition imposed in Eq. (35), the perturbation of the objective function reduces on one hand to

$$\begin{aligned}\delta J(\mathbf{p}) &= -2\mathrm{Re}\sum_{\theta,\omega}\int_0^L \{2Ru\,\delta A^+ + (1+R^2)u\,\delta A^-\}dx \\ &= -\mathrm{Re}\int_0^L \sum_{\theta,\omega} j\omega\{Z_0(1+R)^2 u\,\delta K_e^{-1} \\ &\quad - Z_0^{-1}(1-R)^2 u\,\delta\rho_e\}dx,\end{aligned} \tag{36}$$

and on the other hand, the same perturbation can be expressed as the inner product

$$\begin{aligned}\delta J(\mathbf{p}) = \int_0^L \Bigg[&\frac{\partial J}{\partial \phi}\delta\phi + \frac{\partial J}{\partial \tau_\infty}\delta\tau_\infty + \frac{\partial J}{\partial \Lambda}\delta\Lambda + \frac{\partial J}{\partial \Lambda'}\delta\Lambda' \\ &+ \frac{\partial J}{\partial R_f}\delta R_f\Bigg]dx.\end{aligned} \tag{37}$$

Therefore, with the help of Eqs. (28) and (30), the identification of (36) with (37) leads to the exact expression of the gradient of the objective function. The components of the gradient in terms of each of the parameters to reconstruct are

$$\begin{aligned}\frac{\partial J}{\partial \phi}(x) = -\mathrm{Re}\sum_{\theta;\omega} j\omega\Bigg\{ &Z_0(1+R)^2 u\frac{\partial K_e^{-1}}{\partial \phi} \\ &- Z_0^{-1}(1-R)^2 u\frac{\partial \rho_e}{\partial \phi}\Bigg\},\end{aligned} \tag{38}$$

$$\begin{aligned}\frac{\partial J}{\partial \tau_\infty}(x) = -\mathrm{Re}\sum_{\theta;\omega} j\omega\Bigg\{ &Z_0(1+R)^2 u\frac{\partial K_e^{-1}}{\partial \tau_\infty} \\ &- Z_0^{-1}(1-R)^2 u\frac{\partial \rho_e}{\partial \tau_\infty}\Bigg\},\end{aligned} \tag{39}$$

$$\frac{\partial J}{\partial \Lambda}(x) = -\mathrm{Re}\sum_{\theta;\omega} -j\omega\Bigg\{ Z_0^{-1}(1-R)^2 u\frac{\partial \rho_e}{\partial \Lambda}\Bigg\}, \tag{40}$$

$$\frac{\partial J}{\partial \Lambda'}(x) = -\mathrm{Re}\sum_{\theta;\omega} j\omega\Bigg\{ Z_0(1+R)^2 u\frac{\partial K_e^{-1}}{\partial \Lambda'}\Bigg\}, \tag{41}$$

$$\begin{aligned}\frac{\partial J}{\partial R_f}(x) = -\mathrm{Re}\sum_{\theta;\omega} j\omega\Bigg\{ &Z_0(1+R)^2 u\frac{\partial K_e^{-1}}{\partial R_f} \\ &- Z_0^{-1}(1-R)^2 u\frac{\partial \rho_e}{\partial R_f}\Bigg\}.\end{aligned} \tag{42}$$

In the above relations, all partial derivatives are functions of the components of the vector $\mathbf{p}=[\phi(x),\tau_\infty(x),\Lambda(x),\Lambda'(x),R_f(x)]^T$ via the definitions of ρ_e and K_e in Eqs. (2) and (3). The available analytical expression of the gradient will help the iterative minimization scheme to converge faster to the optimal depth profiles of the five parameters of interest. Classically, one of the most suitable iterative algorithms for our problem is the conjugate gradient method.[25,27,33–35]

A discussion of the numerical calculations, and a presentation of some reconstructions, are given in the following section.

V. PROFILE RECONSTRUCTION OF *x*-DEPENDENT POROUS MATERIAL PROPERTIES: SIMULATIONS

In this section, the computation of the direct scattering solver from Sec. III and the iterative method chosen to solve the optimization problem from Sec. IV are described. The differences with similar problems in electromagnetism[25] or transmission line theory[27] mainly concern the numerical aspects, particularly the choice of the frequency range and the number of points for discretizing our samples. Physical aspects, such as the specific dispersion and absorption processes and properties of porous materials linked to the relations (2)–(5), in which the five material parameters we deal with are introduced, are also treated.

A. Conjugate gradient method

Due to the fact that the objective function $J(\mathbf{p}(x))$ (24) is not a quadratic form and continuous in $\mathbf{p}(x)$, the optimization problem we deal with is nonlinear. The minimization should give accurate enough results when the analytical gradient of J is available. The classical and widely used conjugate gradient method (CG) is one of the most suitable iterative algorithms[25,33–35] for minimization. For clear and rigorous details about the method, the reader can refer to a tutorial (with provided algorithms).[33] For CG algorithms, as well as other optimization techniques, the reader can refer to Polak[35] and Press *et al.*[34] Herein, we describe the CG algorithm for the reconstruction of a single parameter $p(x)$, e.g., the porosity $\phi(x)$. The reconstruction of the other parameters can be simply derived in closed form. The gradient $\partial J/\partial p$ is denoted by $G(p)$. The implementation is as follows:[25,33,34]

- Step 0: Initial approximation: For all x in $[0,L]$, $p(x)=1=p^{(0)}$.
- Step 1: Set $i=0$; Solve the direct problem for all x in $[0,L]$ using Eq. (13)–(16); Calculate the gradient of J with respect to ϕ with the help of Eq. (38); For all x in $[0,L]$, $G^{(0)}=G(p^{(0)})=\partial J/\partial\phi(p^{(0)})$; Set $D^{(0)}=G^{(0)}$. For any iteration order i, $D^{(i)}$ is the search direction.
- Step 2: The line search consists in computing a scalar λ_i such that

$$\forall x \in [0,L], J(p^{(i)} - \lambda_i D^{(i)}) = \min_{\lambda\in\mathbb{R}}\{J(p^{(i)} - \lambda D^{(i)}); \lambda \geq 0\}. \tag{43}$$

- Step 3: Improve the reconstruction by setting $p^{(i+1)}=p^{(i)}-\lambda_i D^{(i)}$.
- Step 4: Compute a new gradient $G^{(i+1)}=G(p^{(i+1)})$ with the help of Eq. (38).

TABLE I. Values of the porous material at $x=0$.

ϕ	τ_∞	Λ (μm)	Λ' (μm)	R_f (Ns m^{-4})	L (cm)
0.96	1.07	273	672	2843	2

- Step 5: if $G^{(i+1)}<\varepsilon$, with $\varepsilon \ll 1$ the convergence criterion, STOP. Else, if $i>$MaxIter, with MaxIter the maximal number of iterations, STOP. Otherwise, update the direction search with a Polak–Ribière condition:[33,35]

$$D^{(i+1)} = G^{(i+1)} + \gamma_i D^{(i)} \tag{44}$$

with

$$\gamma_i = \max\left(\frac{G^{(i+1)^T} G^{(i+1)} - G^{(i+1)^T} G^{(i)}}{G^{(i)^T} G^{(i)}}, 0\right). \tag{45}$$

Set $i=i+1$ and GO TO Step 2.

The minimization over the parameter λ is easily achieved by means of the Nelder and Mead Simplex method.[34,36] This is a fast way to solve an unconstrained minimization problem with only one or a few variables. In case of the simultaneous reconstruction of more than one parameter, λ becomes a vector with as many components as analytical gradients (or material properties to reconstruct).

B. Numerical settings

The general procedure for carrying out the reconstruction of the material profiles is now described.

The simulations are performed for an acoustic plane wave excitation coming from $x<0$, for several angles of incidence ranging from $\theta=0$ rad to $\theta=\pi/3$ rad, and either for $Z_L=Z_0=\rho_f c_f$ (same fluids on each side of the slab) or $Z_L=\infty$ (rigid wall at $x=L$). The advantage of the rigid wall is to detect a reflected signal with large amplitude despite the strongly absorptive material through which the waves travel two times.

1. Generating profiles and scaling

For each parameter to reconstruct, a depth profile is created. The value of each parameter at the boundary $x=0$ is given in Table I. These values were measured on a porous polyurethane foam.[11] When performing the numerical reconstruction of some of the five parameters along the depth x, the others remain constant.

The parameters in Table I can be classified as: (i) those without dimensions (ϕ and τ_∞), which vary around one, their value in the ambient fluid, (ii) those with dimensions of length (Λ and Λ'), which are of the order of several hundreds of micrometers, and (iii) the last one (R_f), whose value is of the order of one to ten thousand.

In order to optimize the detection of variations in the profiles that are of similar range for each parameter, a scaling can be operated. While the parameters keep their physical values when the cost function is evaluated by solving the direct problem at each iteration, they are scaled while the minimization is running by means of the CG algorithm. When a scaling is applied, some new parameters are defined,

$$\forall x \in [0,L], \quad \begin{cases} \Lambda_S = 10\,000\Lambda, \\ \Lambda'_S = 10\,000\Lambda', \\ R_{f_S} = R_f/1000. \end{cases} \tag{46}$$

2. Generating synthetic data

Once the profiles are selected for numerical tests, the effective density ρ_e and K_e are calculated using Eqs. (2) and (3). From the effective characteristic impedance $Z_e=\sqrt{\rho_e K_e}$, the synthetic reflection coefficient R_m is generated.

In a multilayer case, the reflection coefficients $R_m(\omega;\theta)$ are calculated by a recursive method such as the determination of local surface impedances[7] $Z(x)$ in an imbedding geometry,

$$Z(x_i) = \bar{Z}_e(x_i)\frac{-jZ(x_{i+1})\mathrm{cotan}(kdx) + \bar{Z}_e(x_i)}{Z(x_{i+1}) - j\bar{Z}_e\mathrm{cotan}(kdx)}, \tag{47}$$

$$R_m = R(x_1) = \frac{Z(x_1) - \bar{Z}_0}{Z(x_1) + \bar{Z}_0} \tag{48}$$

wherein x_i, $i=1\ldots N_x-1$ is the discretization of x from $x_1=0$ to $x_{N_x}=L$, $dx=x_{i+1}-x_i$ the thickness of an infinitesimal layer, $k=\omega/\bar{c}_e(x,\omega)$ the component of the wave vector along the x axis inside the porous slab in which $\bar{c}_e=\rho_e/\bar{Z}_e$ with $\bar{Z}_e$ from Eq. (22); and $Z(x_{N_x})=\bar{Z}_L$ is the known impedance at $x=L$. Another advantage of generating data by these methods is that the inverse problem can be solved without adding artificial noise. Indeed, the direct solver of the optimization process uses a completely different technique from the solver for generating synthetic data, and thus the "inverse crime"[29] is avoided.

For other inhomogeneous profiles, continuous or not, the synthetic data R_m can be generated by a wave splitting and transmission Green's functions method[9,10] (WS-TGF). The latter is a direct solver that accurately predicts the reflected and transmitted pressure fields, with explicit expressions of the reflection and transmission coefficients, for all the considered frequencies. The WS-TGF method is therefore suitable to generate data in the present work. This method is not similar to the direct solver, except that it uses forward- and backward-propagating waves, e.g., the wave splitting transformation (WS). As explained in Lesselier and Tabbara,[26] the invariant imbedding[21–23,25,27] technique is similar to layer stripping[37] and both are based on iterative resolutions of a Riccati equation as in Eq. (13), whereas the transmission Green's functions (TGFs) technique are linked to the Cholesky factorization of the matrix equation as in (9), which is also similar to the so-called downward continuation.[26,38] Despite these differences between the method of generating data and calculating the field during the optimization, Gaussian noise is added to the synthetic data to simulate experimental error.

TABLE II. Initial values for the reconstructions.

ϕ	τ_∞	Λ (μm)	Λ' (μm)	R_f (Ns m^{-4})
1	1	100	300	1000

3. Initial guess of the parameters

The optimization approach consists of retrieving the material parameter profiles by minimizing the objective function $J(\mathbf{p}(x))$ from Eq. (24). Unfortunately, in general, we do not know *a priori* much about the sample. Then, initial values must be chosen with respect to some general knowledge about rigid frame porous media. In such materials, the values of the porosity $\phi(x)$ and the tortuosity τ_∞ are close to their values in the surrounding fluid, which is one. As to the characteristic lengths, in plastic foams entering into the framework of the rigid frame assumption, we can assume initially that $\Lambda'(x) \approx 3\Lambda(x)$, with $\Lambda(x)$ varying potentially around 100 μm. The flow resistivity $R_f(x)$ has a wide range of variation, but to be able to use the equivalent fluid model, this parameter must not be too large. It is assumed here that $R_f(x)$ can vary from 1000 N s m^{-4} to 20 000 N s m^{-4}. On account of these remarks, but in somewhat arbitrary fashion, we chose the initial values in Table II for the profile reconstructions.

In practice, the instantaneous reflection at the first interface can provide the material parameter values at $x=0^+$. This corresponds to the first pressure wave recorded in reflection in the time domain, due to the sudden addition of the solid frame at the interface at $x=0$. This method is valid as concerns the porosity and the tortuosity in the high frequency asymptotic model.[7,18,20] Then, the values so obtained can be employed as more suitable initial guesses.

To bound the parameter values into the minimizer is similar to add constraints. However, the unconstrained minimization scheme can be kept when an appropriate nonlinear transformation is used. The latter is given by the relation $p_i = f(\psi_i, p_{\min_i}, p_{\max_i})$, $i=1\ldots5$, such that p_i is the ith component of the vector $\mathbf{p}$, i.e., one of the five acoustical parameters of the porous medium, with $p_{\min_i}$ and $p_{\max_i}$ the related lower and upper bounds. The parameter ψ_i is the transformed unconstrained variable on which is applied the conjugate gradient algorithm. An infinite number of nonlinear transformations exists[39] and one is selected and detailed in Appendix B. The use of bounds can help to retrieve the profiles of the characteristic lengths $\Lambda(x)$ and $\Lambda'(x)$ simultaneously with one or several other parameters since only their squared values appear in the equivalent fluid model, in Eqs. (1)–(5). The bounds force the minimizer to seek positive values within a physical meaning.

4. Choice of the weighting function $W_R(\omega)$ and filtering

The weighting function $W_R(\omega)$ is chosen to emphasize the high frequency components in the minimization procedure because these components can reproduce fast variations of the material properties (i.e., because the wavelengths are small). Unfortunately, reconstructions based only on high frequency components are not accurate enough because of unwanted variations in the retrieved profiles.[27] Moreover, the precision of the Runge–Kutta method to solve the differential Eq. (13) decreases when frequency increases. Low frequency components seem to stabilize and smooth the reconstructed profiles. A suitable weighting function is $W_R(\omega) = 100^{-(\omega/\omega_{\max})^2}$ or $W_R(\omega) = 1000^{-(\omega/\omega_{\max})^2}$ if one wants to reduce the high-frequency contribution.

For most of the reconstructions, the band-limited data induce oscillations in the gradient of the objective function, and therefore in the reconstructed profiles. To limit these artifacts, a sliding average window filter is applied to the gradient along the depth x. The length of the window is chosen to be $\texttt{floor}(N_x/N_f)$ with N_x the number of points to discretize the sample thickness, $N_f=128$ the desired number of windows to apply along the thickness and `floor` the floor function that converts any real number a to the highest integer less than or equal to a. The gradient is averaged in this window and then the window is slid from $x=0$ to $x=L$ with a step dx. The result of this procedure is a smoothed gradient and smoothed material parameter profiles, without a noticeable change in the physical information.

5. Convergence criterion

In the given conjugate gradient algorithm, the convergence criterion is based on the value of the gradient of the cost function J. When the minimum of J is found, the gradient should have converged to zero. However, in practice, a small value ε is chosen such that the algorithm stops when the value of the gradient for all x is below the threshold ε. It is difficult to select an optimal criterion ε because the influence of each parameter on the reflection data is different. This means that to screen the variations in the material parameter profiles that do not substantially influence the reflection coefficient a strict criterion is necessary, while a weak criterion is sufficient for the other parameters. To overcome this problem, another convergence criterion is added to the previous one. For all the reconstructions performed in the following, the evolution of the objective function J with the number of iterations N_{iter} is obtained as in Fig. 2. After a certain number of iterations, the cost function J is not decreasing anymore, and its value remains the same for the next iterations. The applied convergence criterion consists of evaluating the variations of $J(\mathbf{p}(x))$ for several consecutive iterations. If the variations of J between five consecutive iterations are below ε, the global minimum might have been found and the algorithm stops. Five iterations were chosen because it can occur that two or three consecutive evaluations of J are equal. For all the reconstructions, the algorithm stops if the convergence criterion on the gradient, normalized by the number of frequency points N_ω, is below $\varepsilon = 10^{-7}$ or 5.10^{-7}, or if the value of J does not vary more than $\varepsilon = 10^{-12}$ after five successive iterations, or if the number of iterations exceeds a given number $N_{\text{MaxIter}}=100$.

With these ingredients, some reconstructions are performed.

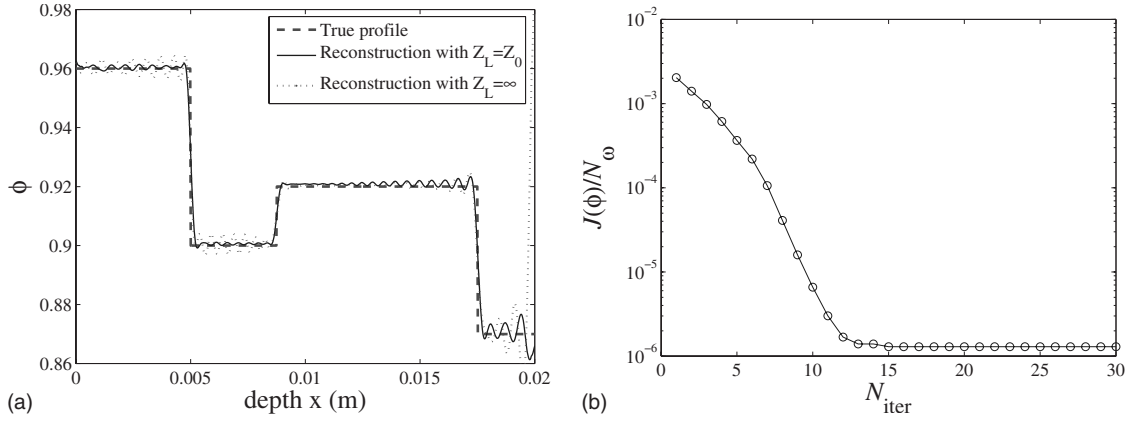


FIG. 2. (Left panel) porosity profile reconstruction after 17 iterations for a two-layered medium surrounded and saturated by air. The computation time was 5 min and 11 s. The data were noiseless. The initial guess is $\phi(x)=1$ for all x in $[0,L]$. (Right panel) decrease of the objective function $J(\phi(x))$ as a function of the number of iterations N_{iter}. The other parameters are constants from Table I.

C. Single profile reconstruction

Each of the five acoustical parameters of a virtual macroscopically inhomogeneous rigid frame porous material are reconstructed independently and successively.

The frequency band [500 Hz, 500 kHz] is discretized into $N_\omega=200$ frequency points with a logarithmic spacing. In practice, according to the value of Λ' from the first layer in Table I, the bandwidth of the incident spectrum signal should not exceed the frequency $f=500$ kHz ($\omega=2\pi f$). This verification is essential to be sure to avoid the diffusion effect that appears when the wavelength of the acoustic excitation is of the order of the average pore radius, which can be related in a rough approximation to Λ'. The lower bound is chosen small enough to have a relatively large frequency range. This can help to improve the reconstruction procedure since the objective function J in Eq. (24) and its gradient of components Eqs. (38)–(42) are summed over a range of frequencies.

The differential Eq. (13) is solved using the fourth order Runge–Kutta method. However, the thickness of the sample must be discretized with a small-enough spatial step dx so that, from $x=0$ to $x=L$, at least $N_x=400$ points are needed in order for the direct solver to converge with enough precision. This choice is a bit empirical but, to allow us to change the frequency band and the material thickness, $N_x=800$ points are used.

As a first test, a four-layered material in which only the porosity is varying is simulated with the help of Eqs. (47) and (48). The optimal profile after 17 iterations and 5 min, 11 s of computation time is presented in Fig. 2 together with the evolution of the objective function J with the number of iterations. The optimization was performed on a pentium 4 PC, with a CPU frequency around 2.99 GHz and 1 Gb of RAM memory.

To test the stability of the algorithm, a larger amount of Gaussian noise was added to the synthetic reflection coefficient. Both the porosity profile and the reflection coefficients are plotted in Fig. 3 for a level of Gaussian noise with a standard deviation set to 0.05. For the three-layer material in panel in Fig. 3(a), the maximum number of iterations was set to 50. One of the reconstruction process stopped after 37 iterations because the last five evaluations of the cost function $J(\phi(x))$ gave the same values. This shows that from one computation to another, there is good stability and that the stopping criteria involved are not always the same from one reconstruction to another. The reconstruction of a continuous profile of the porosity with addition of two air layers is given in panel 3(b) with a filtered contaminated reflection coefficient. The convergence is faster, after 21 iterations and 6 min and 34 s. From these minimization results, one remarks that the reconstructed reflection coefficients are smoother than the "measured" ones. The algorithm therefore appears to be stable with respect to additive Gaussian noise.

The other four parameters, tortuosity, viscous and thermal characteristic lengths and flow resistivity, can also be reconstructed from data pertaining to single-sided normal reflection. However, for a given number of iterations, the accuracy of the reconstructions is not the same from one parameter to another. For instance, the tortuosity is retrieved with an accuracy similar to the one of the porosity, but using the stronger weighting function $W_R=1000^{-(\omega/\omega_{\max})^2}$ and a sliding average window filter to avoid oscillations due to the band-limited data. The use of double-sided reflection data helps to improve the reconstruction, but makes the minimization process slower. Using the same weighting function and filter as the ones used to reconstruct the tortuosity enables the characteristic viscous length Λ is to be well reconstructed with the use of a rigid wall backing. A slightly better reconstruction is obtained by scaling the parameter $\Lambda_S=10\,000\Lambda$. The reconstruction of the characteristic thermal length requires more computational time, but the convergence to the true profile in the studied case is not optimal when single-sided reflection data are employed. With a rigid wall backing, the minimization converged only for plane wave excitation at oblique incidence. The difficulty of reconstructing the characteristic thermal length might come from the fact that thermal losses are much less important than viscous losses in most of the porous materials in which the rigid frame assumption is valid. These losses were even neglected for a long time in the theory of wave propagation in porous materials.[1–5] The fifth parameter, the flow resistivity, is also difficult to reconstruct despite the fact that the objec-

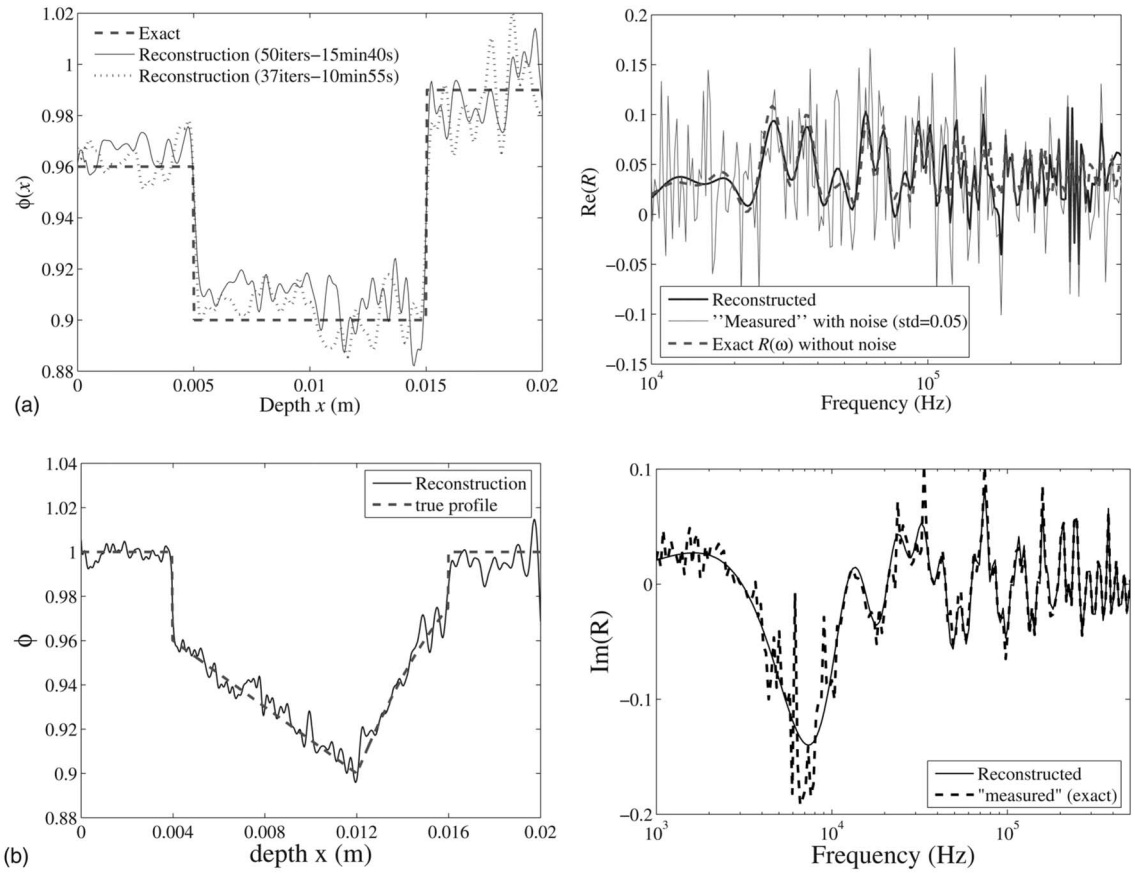


FIG. 3. Convergence and stability of the conjugate gradient method as a function of the noise level. Each row shows the reconstructed profile of the porosity and the corresponding real part of the reflection coefficient. The maximum number of iterations was set to 50. The added noise is Gaussian, with a standard deviation 0.05. The noise added to the reflection coefficient is not filtered in panel (a) and filtered in panel (b). The other parameters are constants from Table I.

tive function is substantially diminished after a few iterations. This result is mainly due to the poor influence of the flow resistivity on the reflection data. This parameter mostly influences low frequency wave propagation attenuation. A more favorable reconstruction can be carried out when the lower bound of the frequency band is decreased.

Examples of reconstructed profiles for the tortuosity, the viscous and thermal characteristic lengths, and the flow resistivity are given in Fig. 4. Results and details of the computational choices are gathered in Table III. According to these results, the combined use of scaling and no backwall ($Z_L=Z_0$) at normal incidence seems appropriate to reconstruct several parameters, except R_f.

The reconstructed profiles of properties that are discontinuous exhibit large jumps, which amount to singularities that are the signatures of numerical difficulties. Continuous parameter profiles, even with discontinuous gradients, induce a less ill-posed optimization problem allowing successful and fast reconstructions. To illustrate the last remark, some examples of single continuous profile (inside the porous slab) reconstructions are presented in Fig. 5. These profiles, which are constructed in a manner similar to that of the porosity in Fig. 3(b), yield more satisfactory reconstructions that are easier to carry out than the ones in Table III and Fig. 4, as we expected. We also studied the case of two air layers, corresponding to the existence of two jumps in the material parameter profiles. These jumps were found to be easier to retrieve than the ones in the profiles of Fig. 4 because air is a homogeneous fluid whose physical properties are known. They are introduced in the computation as known parameters. The reflection coefficients were generated with the help of the wave splitting and transmission Green's functions method (WS-TGF).[9,10] Since this method is considered too similar to the one (WS-II) of the direct problem solver used in the minimization process, a small random Gaussian noise with standard deviation of 0.02 was added to the synthetic data. This corresponds to a large signal to noise ratio (SNR) of 100.

A final remark about the reconstruction of a single parameter: The use of several angles of incidence θ, and of both backings, increases the available data and therefore the rate of convergence and the accuracy of the reconstructions. Several attempts were made to recover $\Lambda'(x)$ (which is difficult to reconstruct) without scaling, for two or three angles of incidence $\theta=0$, $\theta=\pi/6$ and $\theta=\pi/3$. The results are shown in Fig. 6.

D. Simultaneous reconstruction of two profiles

The reconstruction of several material property profiles simultaneously is more difficult to carry out. Since our aim is

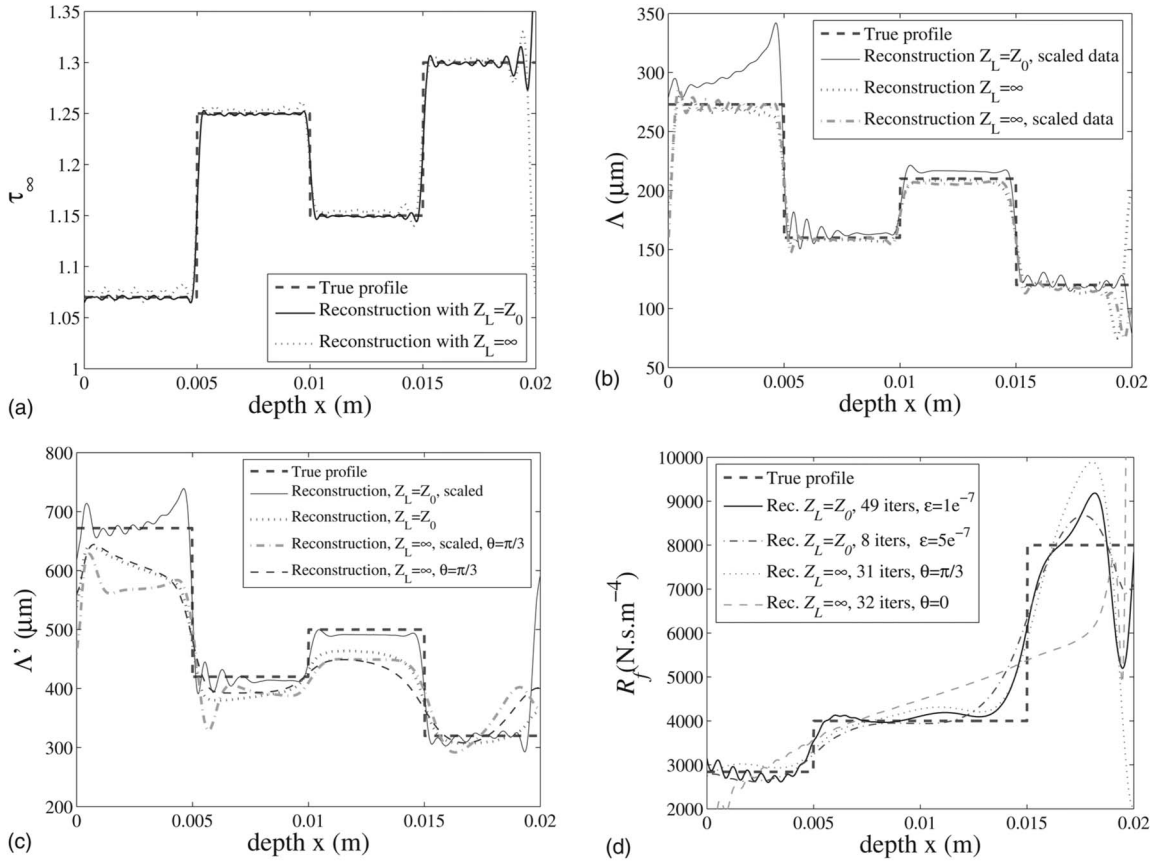


FIG. 4. Reconstruction of (a) the tortuosity τ_∞, (b) the viscous characteristic length Λ, (c) the thermal characteristic length Λ', and (d) the flow resistivity R_f. A small amount of Gaussian noise with a standard deviation of 0.01 is added to the data employed for each reconstruction. The performances of the reconstruction procedures are gathered in Table III. For each reconstruction, the other parameters are constants from Table I.

ideally to retrieve five profiles, several records of reflected signals are necessary. Two very different angles of incidence are chosen, $\theta_1=0$ rad and $\theta_2=\pi/3$ rad.

The porosity and the tortuosity are two parameters without dimensions, that have an important influence on the reflection coefficient. For continuous profiles, the simultaneous reconstruction of these two parameters is quite satisfactory, as depicted in Fig. 7. The reconstruction is obtained after 41 iterations and a bit more than 1/2 h of computation time. The backing is such that $Z_L=Z_0$.

Attempts of simultaneous reconstruction of other couples of material parameters were also made. It was

TABLE III. Convergence of the reconstruction procedure as a function of the number of iterations N_{iter} and computation time. Simulations are performed with or without a rigid wall backing, with or without scaling, and at normal or oblique incidence. `N.A.` stands for Non Accurate, when reconstructions oscillate too much, or when there are giant values at jump positions. `Fails` designates reconstructions that do not vary much from the initial guess. "…" means that no computation was done.

	$Z_L=Z_0$		$Z_L=\infty$		Scaling	θ	ε
	N_{iter}	CPU time	N_{iter}	CPU time		(rad)	
$\phi(x)$	20	6 min 31 s	13	4 min 37 s	No	0	$5e^{-7}$
$\tau_\infty(x)$	24	5 min 31 s	22	8 min 11 s	No	0	$5e^{-7}$
$\Lambda(x)$	58	91 min 26 s	20	43 min 12 s	Yes	0	$5e^{-7}$
	N.A.	N.A.	19	19 min 52 s	No	0	$5e^{-7}$
$\Lambda'(x)$	21	67 min 38 s	N.A.	N.A.	Yes	0	$1e^{-7}$
	6	2 min 38 s	N.A.	N.A.	No	0	$5e^{-7}$
	…	…	12	23 min13 s	Yes	$\pi/3$	$5e^{-7}$
	…	…	11	9 min 53 s	No	$\pi/3$	$5e^{-7}$
$R_f(x)$	Fails	…	Fails	…	Yes	0	$5e^{-7}$
	8	13 min 26 s	32	91 min 40 s	No	0	$5e^{-7}$
	49	96 min 41 s	…	…	No	0	$1e^{-7}$
	…	…	31	99 min 44 s	No	$\pi/3$	$5e^{-7}$

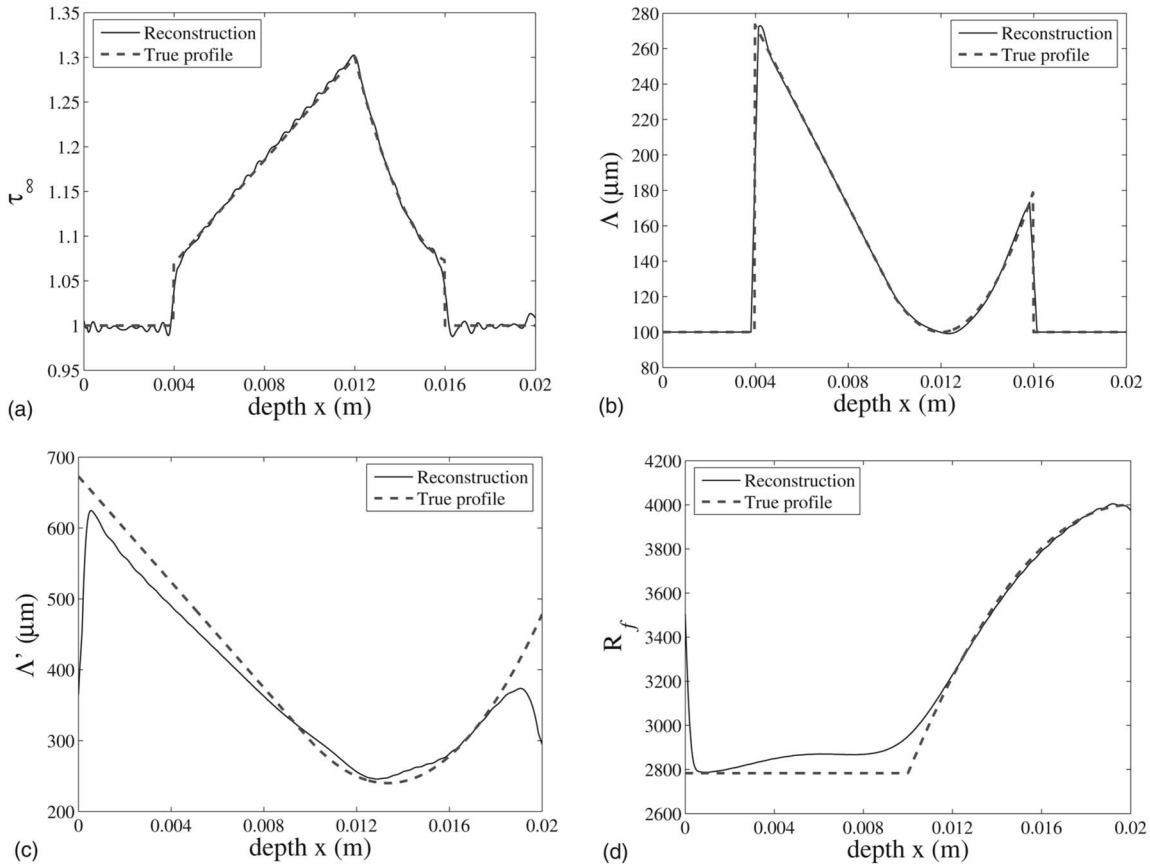


FIG. 5. Reconstruction of (a) the tortuosity τ_∞ after 30 iterations and 10 min, 44 s of computation time, (b) the viscous characteristic length Λ after 30 iterations and 26 min, 41 s, (c) the thermal characteristic length Λ' after 80 iterations and 48 min, 55 s, and (d) the flow resistivity R_f after 30 iterations and 55 min, 11 s. For each reconstruction, a small amount of Gaussian noise with a standard deviation of 0.02 is added to the data. For each reconstruction, the other parameters are constants from Table I.

pointed out previously that the characteristic lengths Λ and Λ' require more computation time to sufficiently diminish the cost function J. Moreover, as noticed for the reconstruction of the porosity-tortuosity couple, the computation time needed to retrieve several profiles simultaneously increases drastically. The way to cope with this problem is to decrease the number of points N_x from 800 to 400 for discretizing the material thickness, and decreasing the higher bound of the frequency range. Thus, only 150 frequency points with a logarithmic spacing are chosen to describe the frequency band [400 Hz, 400 kHz]. In Fig. 8 is presented the simultaneous reconstruction of the porosity ϕ and the viscous characteristic length Λ when the medium is backed by a rigid wall ($Z_L = \infty$).

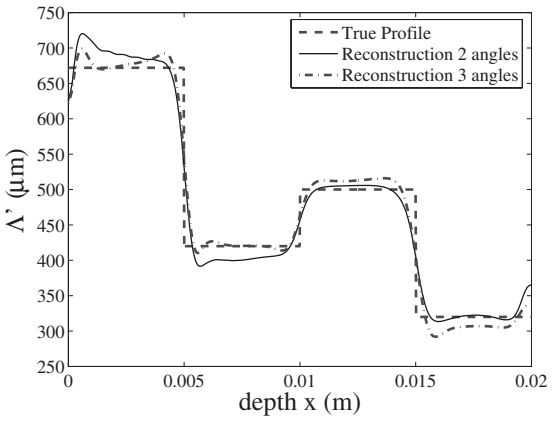


FIG. 6. Reconstruction of $\Lambda'(x)$ with the use of two angles of incidence (θ=0 and $\pi/3$) or three angles of incidence (θ=0, $\pi/6$ and $\pi/3$) with a rigid backing. The number of iterations is respectively 10 and 11, which required 15 min, 2 s and 17 min, 3 s. The data included a small amount of Gaussian noise with a standard deviation of 0.01. The other parameters are constants from Table I.

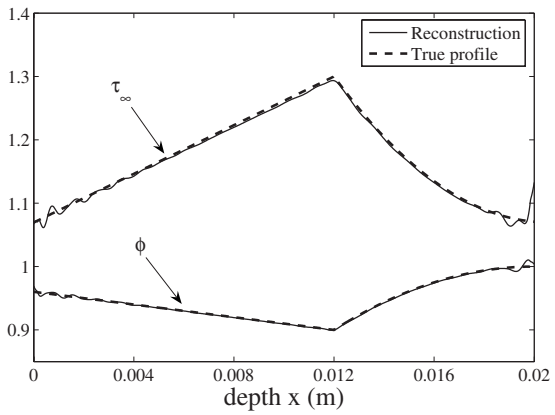


FIG. 7. Simultaneous reconstruction of the porosity and the tortuosity. The number of frequency points is 200, the number of depth points is 800. The reconstruction is given after 40 iterations that took 30 min, 11 s. Addition of a weak Gaussian noise with a standard deviation of 0.01. The other parameters are constants from Table I.

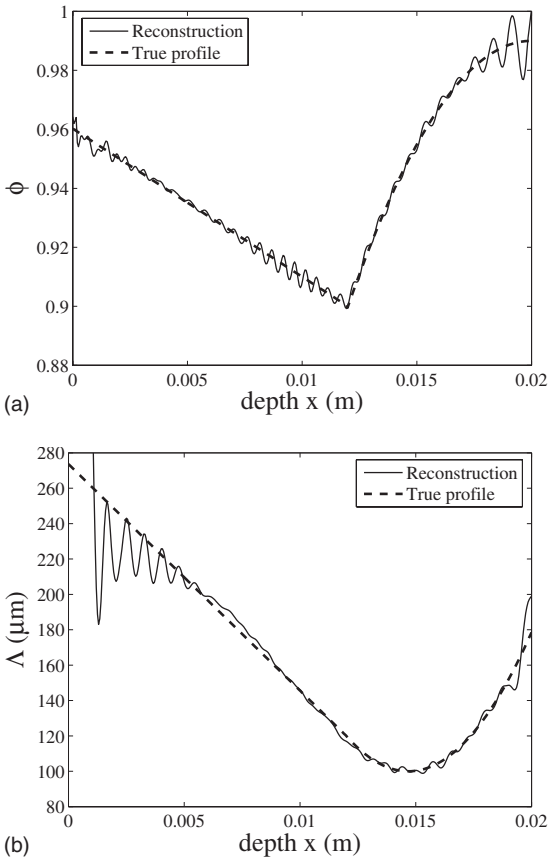


FIG. 8. Simultaneous reconstruction of the porosity ϕ and the characteristic viscous length Λ after 100. The computation time was 55 min and 11 s. The number of points was reduced and the upper bound of the frequency band set to 400 kHz. A small amount of Gaussian noise with a standard deviation of 0.01 was added to the synthetic data. The other parameters are constants from Table I.

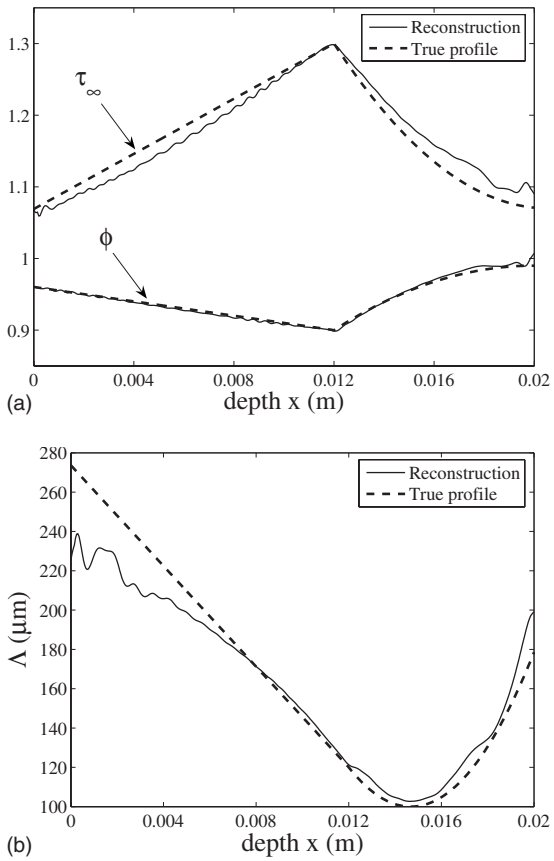


FIG. 9. Simultaneous reconstruction of the porosity ϕ, tortuosity τ_∞ and characteristic length Λ after 100 iterations that took 118 min and 19 s. 150 frequency points and 400 equally spaced points were employed in a frequency band [500 Hz, 500 kHz]. A small amount of Gaussian noise with a standard deviation of 0.01 was added to the synthetic data. The other parameters are constants from Table I.

Satisfactory reconstructions can be obtained in a reasonable amount of time with reduced sets of points to discretize the frequency range as well as the material thickness. However, the optimization procedure is sensitive to these discretizations. The initial guesses of the minimizer are also very important to accelerate the convergence. The reconstructions are faster when they are closer to their true values.

E. Simultaneous reconstruction of more than two material parameter profiles

Employing the reflection data from acoustical excitations at only three different angles of incidence, $\theta_1=0$ rad, $\theta_2=\pi/6$ rad and $\theta_3=\pi/3$ rad, a rigid wall at $x=L$ and a wide-enough frequency band [400 Hz, 400 kHz], only 150 frequency points and 400 equally spaced points along the material thickness are needed to reconstruct three profiles simultaneously, as is shown in Fig. 9. The chosen parameters are the porosity, tortuosity, and viscous characteristic length. They are the most sensitive parameters with respect to the reflection coefficient. Moreover, they are also the parameters of interest in the high frequency asymptotic model for rigid frame porous plastic foams.[7,18,20] The flow resistivity $R_f(x)$ has such a small influence that it vanishes, and the thermal characteristic length $\Lambda'(x)$ is chosen such that $\Lambda'(x)=3\Lambda(x)$ or $\Lambda'(x)=2\Lambda(x)$ according to the foam.

The computation time exceeds 1 h, but is reasonable for such a type of optimization. The results are presented after 100 iterations. Nevertheless, the main variations of the material parameters are already visible after 50 iterations. With experimental data, the expected precision of the reconstructions can be expected to be much less satisfactory.

The result of an attempt to simultaneously reconstruct the same three parameters for a material with strong discontinuities, i.e., a three-layered material, is depicted in Fig. 10. The reconstruction is carried out by minimizing data pertaining to four reflection coefficients. The "measured" ones are obtained by the recursive method from Eqs. (47) and (48). Three of them are simulated without backwall ($Z_L=Z_0$) at three different angles of incidence ($\theta=\pi/6$ rad, $\pi/4$ rad and $\pi/3$ rad) and the fourth one is simulated with a backwall ($Z_L=\infty$) at the angle $\theta=\pi/4$ rad. The frequency band is [1 kHz, 400 kHz], discretized into $N_\omega=300$ equispaced points. The depth is discretized into $N_x=400$ points. The reconstruction took 5 h, 49 min, 47 s and 105 iterations. The result is a bit less accurate and is more difficult to carry out than with a continuous profile. There is a good enough reproductibility but not as satisfactory as with continuous profiles (due to the directions taken by the CG iterations).

The reconstruction of more than three profiles simultaneously can be achieved in a similar way. To improve con-

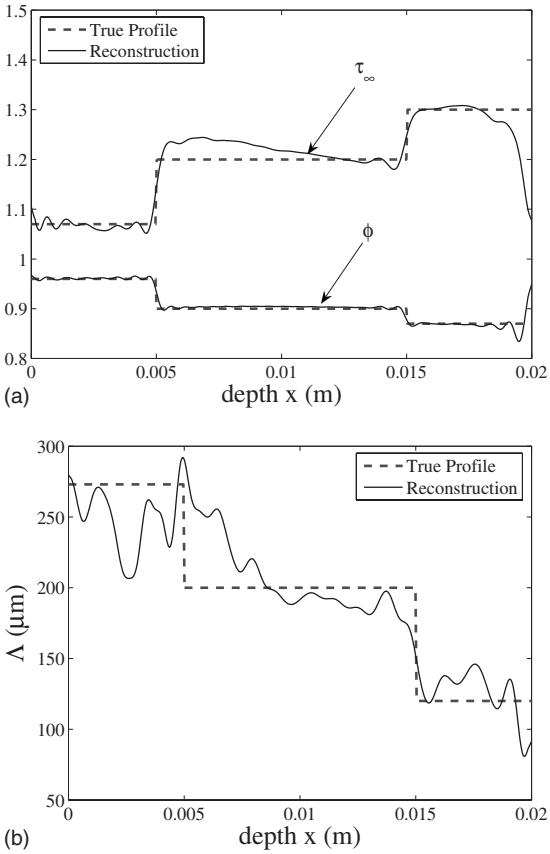


FIG. 10. Simultaneous reconstruction of the porosity ϕ, tortuosity τ_∞ and characteristic length Λ after 105 iterations that took 5 h, 49 min. and 47 s; 300 frequency points and 400 equally spaced points were used in a frequency band [1 kHz, 400 kHz]. The data were noiseless. The other parameters are constants from Table I.

vergence, one can use more sets of data than the number of parameters we aim to reconstruct. For instance, to satisfactorily reconstruct four profiles simultaneously, it can happen that four reflections coefficient spectra at four different angles of incidence are not sufficient. One can then employ data relative to more than four angles of incidence, or one can add measurements with air for all $x>L$ instead of a rigid wall, or one can also impose a rigid wall at $x=0$ while acoustically exciting the sample at $x=L$.

VI. CONCLUSION

An optimization approach to the frequency domain inverse problem is described in this paper in order to reconstruct simultaneously the profiles of several material parameters such as the porosity, the tortuosity, the viscous and thermal characteristic lengths, and the flow resistivity in any unidimensional rigid frame macroscopically inhomogeneous porous materials.

The method is based on the minimization of a suitable objective function, chosen such that its gradient can be analytically obtained. From this amount of information the direct problem is solved iteratively by replacing at each step the values of the parameters to reconstruct by their optimized values. The direct solver is based on the wave splitting and invariant imbedding methods. The optimization approach is based on the conjugate gradient method. Reconstructions of one, two, and even three profiles are obtained with accuracy after only few iterations. The computation time is reasonable in most of the cases. The simultaneous reconstruction of the five profiles is also discussed. The use of several impedances at one end point ($x=L$) and the use of several angles of incidence helps to carry out simultaneous reconstructions based on the reflection data only.

Better optimized results might be obtained when a preconditioner M is used.[33] The latter is a diagonal matrix whose components are those of the main diagonal of the Hessian matrix $H=(\partial^2 J(\mathbf{p}(x))/\partial p_i \partial p_j)$, where p_i, $i=1\ldots5$, refers to the five material parameters to reconstruct. When a preconditioner is used, the direction search $D^{(i)}$ in the conjugate gradient algorithm is replaced by $MD^{(i)}$. When the signal to noise ratio is weak and the variations in the properties to reconstruct are important (i.e., jumps), the optimization problem can become highly ill posed and regularization methods such as Tikhonov regularization[40] might be applied.

In a further work, some questions about the reconstruction of material profiles from measured data need to be solved. For instance, the frequency bandwidths of most of the transducers are not as large as the frequency band used to reconstruct profiles in this article. Therefore, in order to apply the technique detailed in the present paper, numerical methods to extract the necessary large band reflection coefficient are under study.

ACKNOWLEDGMENT

The authors are grateful to Professor M. Norgren for his advice about the present optimization approach.

APPENDIX A: ITERATIVE SOLUTIONS OF LINEAR FIRST ORDER DIFFERENTIAL EQUATIONS

While the Riccati Eq. (13) is solved with a Runge–Kutta method, the other first order differential Eqs. (14), (35) are linear and then easier and faster to solve. An analytical solution of Eq. (14) is easy to carry out. Nevertheless, it may be not the faster way to compute the solution, because for each x, an integral is evaluated. A fast and easy to compute iterative method extracted from the analytical solution is preferred.[28] The technique is general and just a detailed example is given for Eq. (14). The equation is separated as follows in order to have only dx on the right hand side:

$$\frac{dT}{(A^+ + A^- R)T} = dx. \tag{A1}$$

Integrating from x to $x+dx$, and denoting $x_i=x$, $x+dx=x_{i+1}$, $f(x)=f_i$ and $f(x+dx)=f_{i+1}$, with f any function of x, one gets

$$\int_{T_i}^{T_{i+1}} \frac{dT}{(A_i^+ + A_i^- R_i)T} = \int_{x_i}^{x_{i+1}} dx, \tag{A2}$$

$$[\ln T]_{T_i}^{T_{i+1}} = (A_i^+ + A_i^- R_i)\Delta x, \tag{A3}$$

$$T_i = T_{i+1} e^{-(A_i^+ + A_i^- R_i)\Delta x}, \tag{A4}$$

where $\Delta x = x_{i+1} - x_i$.

The recurrence relation to solve Eq. (35) is

$$u_{i+1} = u_i e^{-2(A_i^+ + A_i^- R_i)\Delta x}. \tag{A5}$$

APPENDIX B: NONLINEAR TRANSFORMATION TO ADD BOUNDS TO THE UNCONSTRAINED MINIMIZER

Instead of performing an unconstrained minimization, the addition of bounds on the model parameters may help the optimizer to converge faster to the desired profiles. The bounds in the present case must at least force the algorithm to seek positive real values for each component of the vector $\mathbf{p}(x)$. This gives a lower bound vector $\mathbf{p}_{\min}$. Higher bounds, in $\mathbf{p}_{\max}$, can be guessed according to the physics of porous materials in the rigid frame assumption. However, they should be taken large enough to ensure the accounting of unexpected values to inhomogeneities.

Calling p_i, $i=1,\ldots,5$, the ith component of the vector $\mathbf{p}(x)=[\phi(x),\tau_\infty(x),\Lambda(x),\Lambda'(x),R_f(x)]^T$, an unconstrained vector of components ψ_i in $]-\infty,\infty[$ is defined such that

$$p_i = f(\psi_i, p_{\min_i}, p_{\max_i}) = \frac{p_{\max_i} + p_{\min_i}}{2} + \frac{p_{\max_i} - p_{\min_i}}{2}\sin(\psi_i), \tag{B1}$$

with

$$p_i \to p_{\min_i}, \quad \text{as} \quad \sin(\psi_i) \to -1,$$
$$p_i \to p_{\max_i}, \quad \text{as} \quad \sin(\psi_i) \to +1.$$

The conjugate gradient algorithm is applied to the transformed unconstrained vector ψ. At the kth iteration, the real values of the parameters must be used int the model, and

$$p_i^{(k+1)} = \frac{p_{\max_i} + p_{\min_i}}{2} + \frac{p_{\max_i} - p_{\min_i}}{2}\sin(\psi_i^{(k)} - \lambda_k D_i^{(k)}), \tag{B2}$$

with λ_k a positive real value minimizing the objective function $J(\mathbf{p})$, the gradient of which, in terms of ψ_i, is linked to the search direction $D_i^{(k)}$. Straightforwardly, the components of the gradient are

$$\frac{\partial J}{\partial \psi_i} = \sqrt{(p_{\max_i} - p_i)(p_i - p_{\min_i})}\frac{\partial J}{\partial p_i}. \tag{B3}$$

Remark: In Eq. (B3), it is obvious that the initial guess cannot be set equal to one of the bound. This avoids a nondesired null gradient that implies errors in the Polak–Ribière direction update (45) in the CG algorithm.

[1] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid - I. low frequency range and II. high frequency range." J. Acoust. Soc. Am. **28**, 168–178 (1956); and **28**, 179–191 (1956).
[2] M. A. Biot and D. G. Willis, "The elastic coefficients of the theory of consolidation." J. Appl. Mech. **24**, 594–601 (1957).
[3] M. A. Biot, "Mechanics of deformation and acoustic propagation in porous media." J. Appl. Phys. **33**, 1482–1498 (1962).
[4] M. A. Biot, "Generalized theory of acoustic propagation in porous dissipative media." J. Acoust. Soc. Am. **34**, 1254–1264 (1962).
[5] D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media." J. Fluid Mech. **176**, 379–402 (1987).
[6] Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media." J. Appl. Phys. **70**, 1975–1979 (1991).
[7] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman and Hall, London, 1993).
[8] K. Attenborough, "Acoustical characteristics of porous materials." Phys. Rep. **82**, 179–227 (1982).
[9] L. De Ryck, J. P. Groby, P. Leclaire, W. Lauriks, A. Wirgin, Z. E. A. Fellah, and C. Depollier, "Acoustic wave propagation in a macroscopically inhomogeneous porous medium saturated by a fluid." Appl. Phys. Lett. **90**, 181901 (2007).
[10] L. De Ryck, W. Lauriks, Z. E. A. Fellah, A. Wirgin, J. P. Groby, P. Leclaire, and C. Depollier, "Acoustic wave propagation and internal fields in rigid frame macroscopically inhomogeneous porous media." J. Appl. Phys. **102**, 024910 (2007).
[11] P. Leclaire, L. Kelders, W. Lauriks, N. R. Brown, M. Melon, and B. Castagnède, "Determination of the viscous and thermal characteristic lengths of plastic foams by ultrasonic measurements in helium and air." J. Appl. Phys. **80**, 2009–2012 (1996).
[12] J. P. Groby, L. De Ryck, P. Leclaire, A. Wirgin, W. Lauriks, R. P. Gilbert, and Y. S. Xu, "Use of specific Green's functions for solving direct problems involving a heterogeneous rigid frame porous medium slab solicited by acoustic waves." Math. Methods Appl. Sci. **30**, 91–122 (2007).
[13] O. Forslund and S. He, "Electomagnetic scattering from an inhomogeneous grating using a wave splitting approach," Prog. Electromagn. Waves **19**, 147–171 (1998).
[14] L. L. Beranek, "Acoustic impedance of porous materials," J. Acoust. Soc. Am. **13**, 248–260 (1942).
[15] R. L. Brown and R. H. Bolt, "The measurement of flow resistance of porous acoustic materials," J. Acoust. Soc. Am. **13**, 337–344 (1942).
[16] J.-F. Allard, B. Castagnede, M. Henry, and W. Lauriks, "Evaluation of tortuosity in acoustic porous materials saturated by air," Rev. Sci. Instrum. **65**, 754–755 (1994).
[17] N. R. Brown, M. Melon, B. Castagnede, V. Montembault, W. Lauriks, and P. Leclaire, "Evaluation of the viscous characteristic length of air-saturated porous materials from the ultrasonic dispersion curve," C. R. Acad. Sci. URSS **322**, 121–127 (1996).
[18] Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristegui, and J-Y. Chapelon, "Measuring the porosity and the tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **114**, 2424–2432 (2003).
[19] O. Umnova, K. Attenborough, H. C. Shin, and A. Cummings, "Deduction of tortuosity and porosity from reflection and transmission measurements on thick samples of porous materials," Appl. Acoust. **66**, 607–624 (2005).
[20] Z. E. A. Fellah, C. Depollier, S. Berger, W. Lauriks, C. Aristegui, P. Trompette, and J-Y. Chapelon, "Determination of transport parameters in air-saturated porous materials via reflected ultrasonic waves," J. Acoust. Soc. Am. **114**, 2561–2569 (2003).
[21] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. i. scattering operators and ii. simultaneous reconstruction of dissipation and phase velocity profiles," J. Math. Phys. **27**, 1667–1682 and **27**, 1683–1693 (1986).
[22] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. iii. scattering operators in the presence of a phase velocity mismatch," J. Math. Phys. **28**, 360–370 (1987).
[23] G. Kristensson and R. J. Krueger, "Direct and inverse scattering in the time domain for a dissipative wave equation. iv. use of phase velocity mismatches to simplify inversions," Inverse Probl. **5**, 375–388 (1989).
[24] J. P. Corones, M. E. Davison, and R. J. Krueger, "Direct and inverse scattering in the time domain via invariant imbedding equations," J. Acoust. Soc. Am. **74**, 1535–1541 (1983).
[25] M. Gustafsson and S. He, "An optimization approach to multi-dimensional time domain acoustic inverse problems," J. Acoust. Soc. Am. **108**, 1548–1556 (2000).
[26] D. Lesselier and W. Tabbara, "Problèmes unidimensionnels de diffraction inverse en acoustique et électromagnétisme. etude bibliographique (Unidimensional inverse scattering problems in acoustic and electromagnetism. Bibliographical study)," J. Acoust. **1**, 363–384 (1988).
[27] M. Norgren and S. He, "An optimization approach to the frequency-domain inverse problem for a nonuniform lcrg transmission line," IEEE Trans. Microwave Theory Tech. **44**, 1503–1507 (1996).
[28] C. MacKay, D. Hayward, S. McKee, A. J. Mulholland, and R. A. Pethrick, "An inverse problem of reconstructing the electrical and geometrical pa-

rameters characterizing airframe structures and connector interfaces," Inverse Probl. Eng. **15**, 177–190 (2007).
[29]A. Wirgin, *The Inverse Crime*. Available at http://hal.archives-ouvertes.fr/hal-00001084/fr/, last viewed online on 04/03/2008.
[30]R. J. Krueger and L. Ochs, "A Green's function approach to the determination of internal fields," Wave Motion **11**, 525–543 (1989).
[31]J. le Rond d'Alembert, *Opuscules Mathématiques ou Mémoires sur Différens Sujets de Géométrie, de Méchanique, d'Optique, d'Astronomie (Mathematical Opus or Manuscripts on Several Subjects About Geometry, Mechanics, Optics and Astronomy)* (Chez David, Paris, 1761), pp. 1761–1780. Available at http://gallica.bnf.fr/ark:/12148/bpt6k62394p/ last viewed online on 7/31/08.
[32]S. He, S. Ström, and V. H. Weston, *Time Domain Wave-Splittings and Inverse Problems* (Oxford University Press, Oxford, 1998).
[33]J. R. Shewchuk, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*. Available at www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf, last viewed online on 04/04/2008.
[34]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[35]E. Polak, *Computational Methods in Optimization* (Academic, New York, London, 1971).
[36]J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. **7**, 308–313 (1965).
[37]A. E. Yagle, "A layer stripping fast algorithm for reconstructing a lossy layered medium," J. Acoust. Soc. Am. **82**, 927–936 (1987).
[38]K. P. Bube and R. Burridge, "The one-dimensional inverse problem of reflection seismology," SIAM Rev. **25**, 497–559 (1983).
[39]T. M. Habashy and A. Abubakar, "A general framework for constraint minimization for the inversion of electromagnetic measurements," Electromagn. Waves **46**, 265–312 (2004).
[40]A. Franchois and C. Pichot, "Microwave imaging-complex permittivity reconstruction with a Levenberg-Marquardt method," IEEE Trans. Antennas Propag. **45**, 203–215 (1997).

Effect of the cochlear microphonic on the limiting frequency of the mammalian ear

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

(Received 11 March 2008; revised 4 June 2008; accepted 8 June 2008)

Electromotility is a basis for cochlear amplifier, which controls the sensitivity of the mammalian ear and contributes to its frequency selectivity. Because it is driven by the receptor potential, its frequency characteristics are determined by the low-pass *RC* filter intrinsic to the cell, which has a corner frequency about 1/10th of the operating frequency. This filter significantly decreases the efficiency of electromotility as an amplifier. The present paper examines a proposal that the cochlear microphonic, the voltage drop across the extracellular medium by the receptor current, contributes to overcome this problem. It is found that this effect can improve frequency dependence. However, this effect alone is too small to enhance the effectiveness of electromotility beyond 10 kHz in the mammalian ear. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2953317]

PACS number(s): 43.64.Bt, 43.64.Ld, 43.64.Nf [BLM] Pages: 1607–1612

I. INTRODUCTION

The sensitivity and frequency selectivity of the mammalian ear depend on the motile activity of outer hair cells (OHCs). The electromotility of OHCs is likely responsible for this function.[1–4] However, its dependence on the receptor potential (RP) may limit its efficiency because the RP is attenuated by the intrinsic low-pass *RC* filter of the cell, the corner frequency of which is significantly lower than the operating frequency. This is known as the *RC* time constant problem.[5,6] To be able to evaluate the significance of this problem, we will need a scale with which force production by electromotility is compared. In an earlier report,[7] we assumed that this motility must counteract viscous drag,[8] specifically the shear drag in the gap between the tectorial membrane and the reticular lamina.[9] We found that this comparison leads to a frequency limit and that the limiting frequency obtained for the ear geometry of chinchillas was about 10 kHz, respectable but still short of covering the auditory frequency. This result suggests the presence of factors that enhances the effectiveness of electromotility.

To address the *RC* time constant problem, several mechanisms that can enhance the efficiency of electromotility have been proposed. Those mechanims include piezoelectric resonance,[7,10] modulation by a stretch-sensitive Cl^- current,[11,12] fast voltage-gated *K* channels,[7,13] and the effect of the cochlear microphonic (CM).[14,15] Of those mechanisms, a previous analysis[7] showed that pure piezoelectric resonance that does not involve the mechanoelectric transducer channel is not effective and that fast voltage-gated *K* channels, which can counteract the capacitive current, is capable of extending the frequency limit beyond the 10 kHz limit. Now it appears possible to extend our analysis to include the effect of the CM on unstimulated outer hair cells[14,15] because recent experiments[16,17] have provided details regarding traveling waves and the CM.

Dallos and Evans[14,15] posited the following. The CM, which is a voltage drop generated by the receptor current of outer hair cells across the extracellular medium, has relatively flat frequency dependence. In addition, unstimulated outer hair cells should act as voltage dividers with relatively flat frequency dependence because at low frequencies their performance is determined by the resistance ratios of their apical and basolateral membranes and at high frequencies the performance is determined by the capacitance ratios.[14,15] Thus the oscillatory membrane potential in unstimulated OHCs induced by the CM has relatively flat frequency dependence, and this potential can be larger than the receptor potential of stimulated OHCs at high frequencies.

Here we examine whether or not the CM can contribute to overcoming the *RC* problem by constructing a simple theoretical model that would set an upper bound of efficiency.

First, we assume that the stimulus level is extremely low and give an analytical expression for the RP, the CM, which is locally produced by the receptor current of OHCs,[18] and the oscillatory potential induced in outer hair cells by the CM. This calculation could be justified for the following reasons. Since the responses of the cochlea to stimuli with very low intensities are extremely localized,[19,20] we estimate the external potential near the locus of the characteristic frequency, assuming that the major current source is local. This assumption, which is rather approximate, appears to be compatible with the value of ~200 μm (Ref. 16) for the wavelength of the traveling wave near the peak and estimates of >300 μm (Ref. 21) and 40 μm[17] for the space constant of the CM.

Second, instead of a detailed model of cochlear mechanics, we assume that mechanical energy generated by the CM is optimally converted into mechanical energy. We further assume that mechanical energy made available from electromotility is optimally used to counteract viscous drag. Then

[a]Electronic mail: ki2s@nih.gov.
[b]Electronic mail: sulb@nidcd.nih.gov.

by considering energy balance we examine whether or not this effect increases the frequency limit that is imposed by viscous drag.

II. EQUIVALENT CIRCUIT FOR HAIR CELLS

Following Dallos and Evans,[15] we consider two kinds of OHCs. One kind includes those cells that are stimulated at their characteristic frequency. These cells have receptor currents, which generate the RP across their plasma membrane and the CM in their vicinity in the extracellular space. Although these cells convert electrical energy made available by the RP into mechanical energy, let us refer them to *sensor* cells for convenience. Another kind includes unstimulated OHCs. These cells have oscillating potential not because of stimulation to their hair bundles but because of the CM. Let us call this potential "pseudo-receptor-potential" (pRP). We refer to those cells as *actuator* cells because their pRP drives those cells. For simplicity, we assume that these two kinds of cells, sensors and actuators, share the same electrical properties that include the apical capacitance C_a, the basal capacitance C_b, the resting apical resistance R_0, and the basal resistance R_K.

Since the resting potential of OHCs is about −70 mV,[22] close to the reversal potential e_K of about 100 mV for K^+,[23] the basal conductance is dominated by K^+ current. The apical membrane is exposed to the endolymph with endocochlear potential e_{ec} of about 80 mV,[24] measured from the perilymph, which surrounds the basolateral membrane.

The circuit is described by

$$I_1 = \frac{V_1 - e_K}{R_K} + C_b \frac{d}{dt} V_1 = \frac{V_e - V_1}{R_a} + C_a \frac{d}{dt}(V_e - V_1), \quad (1)$$

$$I_2 = \frac{V_2 - e_K}{R_K} + C_b \frac{d}{dt} V_2 = \frac{V_e - V_2}{R_0} + C_a \frac{d}{dt}(V_e - V_2), \quad (2)$$

$$I_1 + nI_2 = \frac{e_{ec} - V_e}{R_e}, \quad (3)$$

where V_1 and I_1 are the membrane potential and current of sensor cells, respectively, and V_2 and I_2 are those for actuator cells. The voltage drop $e_{ec} - V_e$ is produced by currents through the external resistance R_e. The factor n is the number ratio of actuator cells to sensor cells.

We placed the extracellular resistance R_e on the apical side of OHC's (i.e., in the scala media) in the equivalent circuit diagram (Fig. 1). However, it is also reasonable to have the extracellular resistance on the basolateral side of the OHCs. These two cases give the same result for the ac voltage drop because the circuit is closed at the ground.

Let us assume that the basilar membrane moves sinusoidally with a small amplitude x at an angular frequency $\omega(=2\pi f)$. The reason for this assumption is that the sensitivity of the basilar membrane is largest for small signals,[20] and therefore examining the case of small signals must be sufficient for examining the RC time constant problem. The resistance R_a of the apical membrane, which responds to the motion of the basilar membrane without a phase delay due to fast gating of mechanotransducer channels, can be described by

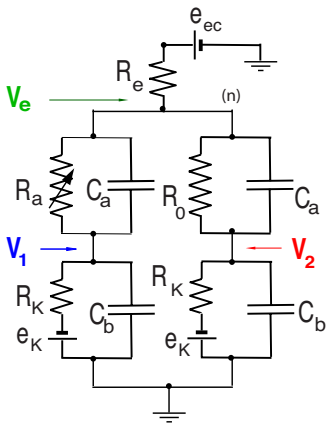


FIG. 1. (Color online) An equivalent electric circuit that represents the configuration in which two kinds of outer hair cells are connected. On the left is the sensor cell, which has mechanotransducer conductance represented by the resistance R_a. The transducer current induces the receptor potential V_1 in the cell as well as the cochlear microphonic V_e across the external resistance R_e. The apical capacitance is C_a; the basolateral membrane has resistance R_b and capacitance C_b. On the right is the actuator cell. The property of the actuator cell is identical to the sensor cell, with the exception that the apical resistance is constant at R_0. The cochlear microphonic induces pseudo-receptor-potential V_2 in the actuator cell.

$$R_a(t) = R_0(1 - \epsilon \exp[i\omega t]), \quad (4)$$

with $\epsilon \ll 1$. The relationship between x and ϵ can be given by using the sensitivity g of the mechanoelectric transducer channel to basilar membrane displacement,

$$\epsilon = gR_0 x. \quad (5)$$

The potentials V_1, V_2, and V_e can be expressed by

$$V_1(t) = V_{10} + e_K + v_1 \exp[i\omega t], \quad (6)$$

$$V_2(t) = V_{20} + e_K + v_2 \exp[i\omega t], \quad (7)$$

$$V_e(t) = V_{e0} + e_K + v_e \exp[i\omega t], \quad (8)$$

where v_1, v_2, and v_e are small complex amplitudes.

Although the membrane capacitance C_b depends on the membrane potential due to the membrane motor, we treat it as constant. Because the voltage changes are small, the voltage-dependent component of the capacitance contributes only to second- and higher-order terms.

A. Perturbation solution

The zeroth-order terms lead to

$$V_{10} = V_{20} = \frac{R_K V_0}{R_0 + R_K + (1+n)R_e}, \quad (9)$$

$$V_{e0} = \frac{(R_0 + R_K)V_0}{R_0 + R_K + (1+n)R_e}, \quad (10)$$

where $V_0 = e_{ec} - e_K$. Here $V_{10}(=V_{20})$ is the resting potential of OHCs. The dc component of the membrane potential of both

types of cells are equal because it is determined by the time independent part of the resistance, which is equal in both.

The first-order terms lead to

$$\frac{v_1}{\epsilon} = i\frac{R_0 V_{10}}{D_n}\{nR_e + R_K(1 + i\omega C_b nR_e) + R_0[1 + i\omega C_b R_K + i\omega C_a(R_K + nR_e + i\omega C_b nR_K R_e)]\}, \quad (11)$$

$$\frac{v_2}{\epsilon} = -\frac{iR_0 R_e V_{10}}{D_n}(1 + i\omega C_a R_0)(1 + i\omega C_b R_K), \quad (12)$$

$$\frac{v_e}{\epsilon} = -\frac{R_0 R_e V_{10}}{R_K D_e}(1 + i\omega C_b R_K), \quad (13)$$

with

$$D_n = -i\{R_K + R_0[1 + i\omega(C_a + C_b)R_K]\}D_e, \quad (14)$$

$$D_e = -(1+n)R_e - R_K[1 + i\omega C_b(1+n)R_e] - R_0(1 + i\omega C_b R_K + i\omega C_a\{(1+n)R_e + R_K[1 + i\omega(1+n)C_b R_b]\}). \quad (15)$$

Here v_1 is the ac component of the RP, v_2 is the ac component of the pRP, and v_e is the CM.

In the special case of $R_e=0$, both CM and pRP disappear, i.e., $v_e=v_2=0$. The RP v_1 of the sensor cell turns into v_0 and

$$\frac{v_0}{\epsilon} = \frac{R_0 V_{10}^{(0)}}{R_0 + R_K + i\omega R_0 R_K(C_a + C_b)}, \quad (16)$$

where $V_{10}^{(0)}$ is V_{10} with $n=0$.

III. MAGNITUDES OF OSCILLATING POTENTIALS

Here we numerically examine the phase and the magnitude of the pRP and compare them with those of the RP. A set of typical parameter values is given in Table I.

Numerical examination shows that the magnitude of the pRP is less than that of the RP if $n \geq 1$ for the entire frequency range. However, the pRP increases with decreasing n, and a reversal takes place at a high frequency range if $n < 1$ (Fig. 2). The reason for the reversal is the phase. The phase of the pRP is similar to that of the CM, which differs in phase from the RP by about π. Namely, the RP suffers destructive interference from the CM.

These observations are consistent with the high frequency asymptotes. For high frequency, Eqs. (11) and (12) yield

$$\frac{v_1}{\epsilon V_{10}} \approx -i\frac{n}{n+1}\frac{1}{\omega(C_a + C_b)}, \quad (17)$$

$$\frac{v_2}{\epsilon V_{10}} \approx i\frac{1}{n+1}\frac{1}{\omega(C_a + C_b)}. \quad (18)$$

These equations show that at high frequencies the inequality $|v_1| > |v_2|$ holds for $n > 1$ and $|v_1| < |v_2|$ for $n < 1$. They also show the phase difference of π.

TABLE I. Parameter values. The values used reflect properties of basal cells.

Quantity	Used	Measured (estimated)
C_a	5 pF	(3 pF)[a]
C_b	10 pF	(10 pF)
C_a+C_b	15 pF	⩾15 pF[f]
R_K	10 MΩ	10 MΩ,[b] 20 M[f]
R_0	60 MΩ	(60 MΩ)[c]
R_e	1 MΩ	(0.7 MΩ)[d]
g	1 nS/nm	1 nS/nm[e]

[a]Estimation based on the membrane area.
[b]Reference 27.
[c]Assumes 50% opening of the maximal hair bundle conductance, which is 9.2 nS in a perilymphatic medium (Ref. 28) multiplied by 3 to account for low Ca^{2+} endolymph (Ref. 29).
[d]The amplitude of cochlear microphonic near outer hair cells is 0.1 mV for basilar membrane motion with 1 nm amplitude (Ref. 17), which elicits 0.15 nA, given the sensitivity of 1 nS/nm (see footnote e) and 150 mV potential drop. This gives 0.7 MΩ for R_e.
[e]Obtained by multiplying the steepest slope for hair bundle open probability 1/(25 nm) multiplied by 28 nS conductance (Refs. 28 and 29), assuming a 1:1 ratio of hair bundle displacement to basilar membrane displacement. This value is also consistent with the peak sensitivity of 0.7 nS/nm (Ref. 30) obtained from the hemicochlea preparation, considering the blocking effect of high Ca^{2+} concentration in the perilymphlike medium.
[f]Reference 5.

IV. COLOCALIZATION OF TWO TYPES OF OHCs

If actuator cells are colocalized with sensor cells, what is expected? Since electromotility is proportional to the oscillating potential of OHCs, the effect of CM can be assessed by the ratio of the weighted mean of the receptor potential and the pRP $|v_1+nv_2|/(1+n)$ to the RP $|v_1|$. Since the phase

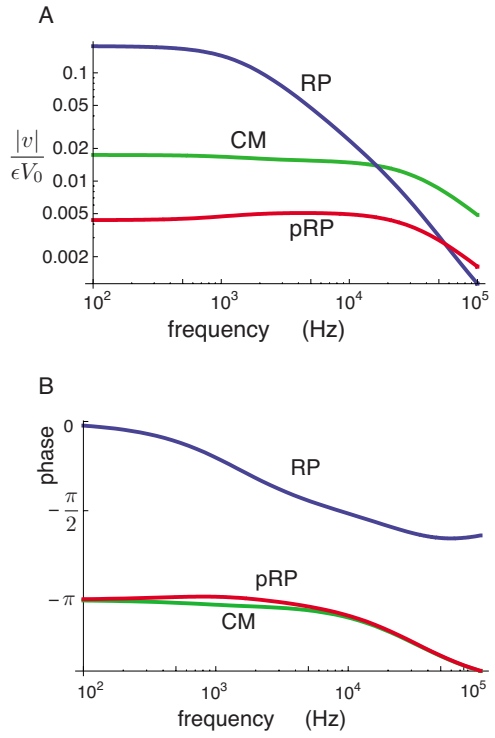


FIG. 2. (Color online) The magnitude (A) and phase (B) of the RP, CM, and pRP. The values of the parameters are shown in Table I. The value for n is 0.5.

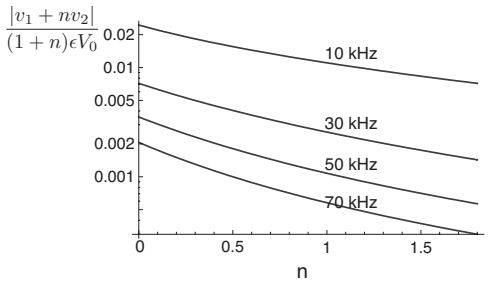


FIG. 3. The dependence of the mean amplitude $(|v_1+nv_2|)/(n+1)$ on the number ratio n of the actuator cells to the sensor cells. The n-dependence of the mean amplitude depends on the frequency f. The top trace is for 10 kHz and the bottom trace is for 70 kHz, with frequency incremented in 20 kHz steps.

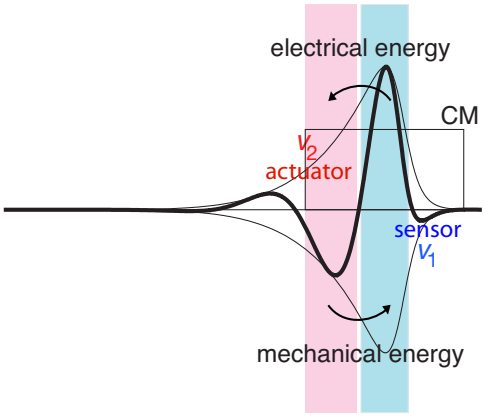


FIG. 4. (Color online) A simplified model showing lateral energy transfer. The bold line represents a traveling wave that propagates from the basal end (left). The sensor cells have the RP V_1 and are in the zone of 1/2 wavelength (the shaded band on the right-hand side) centered at the characteristic location of the frequency, where the envelope (the thin line) peaks. The actuator cells have the pRP V_2 and are at a more basal zone of 1/2 wavelength (the shaded band on the left-hand side) centered at the point where traveling wave's phase is advanced by π. The spread of the CM is represented by the box centered at the sensor cells and extends $3/4\ \left(=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\right)$ wavelength in both directions.

of v_2 differs from that of v_1 approximately by π, the RP and the CM have destructive interference. Thus we expect

$$|v_1 + nv_2|/(1+n) < |v_1|, \tag{19}$$

which is illustrated in Fig. 3. This inequality is obvious for the high frequency asymptotes. For this reason, having both actuator cells in the immediate neighborhood of the sensor cells is not only inconsistent with morphological observations, but also detrimental to the effectiveness of electromotility.

V. UPPER BOUND OF CM UTILIZATION

Following Dallos and Evans,[15] here we assume that the sensor cells are in the characteristic location of frequency ω and the cells that uses the CM are more basal to the characteristic location.

Because we are interested in finding the frequency limit of the ear, we consider energy balance at the most basal part of the cochlea. We examine an upper bound of the efficiency with which actuator cells use the pRP by considering energy balance alone. Specifically, the CM generated by OHCs in the location of peak amplitude is used exclusively by those cells in a slightly more basal location to convert electrical energy into mechanical energy to be sent back to those cells that generate the CM.

Where are of the actuator cells located in this hypothesis? The conditions for optimal location for the actuator cells would be as follows:

(1) The mechanical energy output of the actuator cells depends on cell displacement, which is associated with basilar membrane movement. For this reason, a higher efficiency of actuators in producing mechanical energy requires a larger amplitude of basilar membrane motion at the site of these cells.
(2) Displacements of the basilar membrane generate receptor currents that contribute to the CM. For this reason, the amplitude of basilar membrane motion at the location of the actuator cells must be small enough not to diminish the CM by destructive interference.
(3) The RP of actuator cells must have constructive interference with the CM. The zone of constructive interference is where phase advance is between $\pi/2$ and $3\pi/2$, optimally π. Without constructive interference, all the electrical energy of a cell is spent in counteracting the viscous drag at the location of the cell.
(4) The efficiency of energy transmission decreases with the distance between sensors and actuators. Transmission of electrical energy depends on a space constant, which is $\sim 40\ \mu$m (Ref. 17) or $<300\ \mu$m,[21] perhaps somewhat shorter than the wavelength of about 200 μm.[16]

Condition (3) on the phase requires that the actuator cells be located between 1/4 and 3/4 wavelength basal to the characteristic location of the frequency. The most effective location is between 1/4 and 1/2 wavelength basal, where the amplitude is larger [condition (1)] and closer to the sensors [condition (4)].

To evaluate an upper bound of the efficiency with which actuator cells convert electrical energy associated with the CM, we make the following simplifications (Fig. 4). First, the sensor cells are OHCs in the 1/2 wavelength zone centered at the location of the characteristic frequency. Second, the actuator cells are OHCs in the 1/2 wavelength zone centered at the location 1/2 wavelength more basal to the characteristic location. Third, the CM spreads $3/4\left(=\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\right)$ wavelength in the basal direction from the characteristic location to cover the actuator cells and abruptly diminishes in a step. Because a space constant does not have directionality, the CM should spread in the apical direction symmetrically as it does in the basal direction. These assumptions lead to a value of 2 for the ratio n, which determines the attenuation of the CM. The exact value for n is not critical for our analysis, as will be shown later. Fourth, the displacement amplitude of the actuator is equal to that of sensors. This condition to maximize the actuator output (the condition 1) would reduce the CM [condition (2)]. However, we assume that the CM is

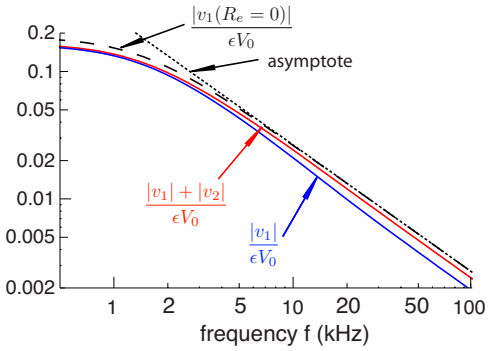


FIG. 5. (Color online) Relative magnitude of electromotility output, which is proportional to the electric potential plotted against the frequency. The combined magnitude $(|v_1|+|v_2|)/(\epsilon V_{10})$ is compared with $|v_1|/(\epsilon V_{10})$ and $|v_0|/(\epsilon V_{10})$ (dashed line), the magnitude of the sensor cell's receptor potential with and without (dashed line) the external resistance R_e, respectively. The asymptotic form of $|v_0|/(\epsilon V_{10})$ for high frequencies $1/[\omega R_b(C_a+C_b)]$ (dotted line) was used in the previous treatment (Ref. 7).

not affected. These assumptions, each of which would be too optimistic, are again intended to give an upper bound of efficiency.

With these assumptions, an upper bound of the energy generated by an OHC per cycle to counteract viscous drag is represented by

$$\bar{E}_{\mathrm{OHC}} = \int_0^{2\pi/\omega} dt\phi \cos^2(\omega t)(|v_1| + |v_2|)x, \tag{20}$$

where ϕ is the force generation per unit voltage change and x is the displacement of the basilar membrane at the characteristic location.

Since the CM is the product of the transducer current of OHCs through the extracellular space, its effect on enhancing electromotility requires introducing an extracellular resistance R_e. However, the RP increases as $R_e \to 0$. As illustrated in Fig. 5, the reduction of the RP from v_0 to v_1 due to the external resistance is somewhat compensated by the effect of utilizing the CM under the most favorable condition. However, the combined amplitude $|v_1|+|v_2|$ still remains less than $|v_0|$, i.e., the RP without external resistance, throughout the frequency range.

With Eqs. (5), (17), and (18), the asymptotic form for high frequencies is

$$\bar{E}_{\mathrm{OHC}} \approx \pi\phi \frac{gR_0V_{10}}{R_K\omega(C_a+C_b)}x^2. \tag{21}$$

VI. LIMITING FREQUENCY REVISITED

In a previous analysis,[7] viscous drag in the gap between the tectorial membrane and the reticular lamina has been addressed. It showed that the drag is proportional to the velocity of shear motion and that the friction coefficient γ is expressed by $\gamma = \eta S/d$, where η is the viscosity of the fluid, S the area of the gap per OHC, and d the gap. The viscous loss per cycle per OHC at the characteristic location is then

$$E_{\mathrm{drag}} = \pi\gamma\omega x^2. \tag{22}$$

The condition for OHC electromotility to counteract viscous drag is $\bar{E}_{\mathrm{OHC}} \geqslant E_{\mathrm{drag}}$. This leads to a condition for the frequency ω,

$$\omega^2 \leqslant \alpha(n) \cdot \frac{\phi g V_0}{g(C_a+C_b)}, \tag{23}$$

with a factor

$$\alpha(n) \equiv \frac{V_{10}R_0}{V_0R_K} = \frac{R_0}{R_0+R_K+(1+n)R_e}, \tag{24}$$

which is smaller than unity.

In the absence of extracellular resistance, the limiting frequency is somewhat higher because $\alpha(0) > \alpha(n)$ for $n > 0$.

Because $\alpha(n) < 1$, inequality (23) gives a frequency limit lower than that given by

$$\omega^2 \leqslant \frac{\phi g V_0}{\gamma(C_a+C_b)}, \tag{25}$$

which was obtained assuming $R_0 \gg R_K$ and $R_e = 0$[7]. With $\phi = 0.1$ nN/mV (Ref. 25) and $\gamma = 1.4 \times 10^{-7}$ N s/m,[7] condition (25) leads to the frequency limit of about 10 kHz for guinea pigs and chinchilla.[7] It should be noted that no single factor has a decisive effect on the frequency limit because the limit depends on the square root of each factor. Although the value for the friction coefficient γ of the subtectorial space, for example, is determined based on electron microscopy data that may need correction for shrinkage, such a correction has a relatively modest effect on the frequency limit because of the square root dependence.

VII. DISCUSSION

Our model is intended to obtain an upper bound of the CM's effect on enhancing electromotility, and it is much simpler than the configuration in which most experimental data are obtained. Specifically, we evaluated the CM generated by the transducer current of OHCs in one location of the cochlea associated with basilar membrane motion with small amplitudes. However, in most experimental conditions, the transducer current that contributes to the CM passes through OHCs in a much wider area, and the contributions of these OHCs are determined by a space constant. This explains why the calculated frequency dependence of the potential (Fig. 2) differs from that of the experimentally observed one.[17]

In an experimental condition where the cochlea is stimulated via stapes and an electrode for recording the CM is placed at a location in the organ of Corti, an increase in the frequency ω with constant stimulation intensity at the stapes shifts the location of maximum amplitude basally and changes the amplitude x of the basilar membrane motion at a given location, where the recording electrode is placed.

If the stimulation frequency is increased from below the best frequency of the recording locus, the peak of the traveling wave moves closer to the recording location from a more apical location. This basal shift of the current source advances the phase of the CM closer to π, from 130° to 164°in Fridberger *et al.*,[17] as the frequency approaches the best frequency. The phase of CM further advances when the

frequency further increases beyond the best frequency for stimulation intensities below 40 dB sound pressure level.[17] This explanation is consistent with the interpretation given by Fridberger *et al.*[17]

Our sensor cells are at the location of the highest characteristic frequency, and we are only interested in high frequency response. We do not have to consider basal shifts of the current source as frequency increases. In addition, the phase of the CM experimentally observed near the best frequency is close to π,[17] consistent with our model. For these reasons, our model does not contradict the experimental data for our purpose. The phase of CM only decreases with increasing frequency, reflecting the *RC* filter of the circuit.

Finally, our result that extracellular potential can improve the effectiveness of electromotility is consistent with a recent model evaluation[26] and an experimental study.[17] Our result is further consistent with a conclusion of the latter[17] in that this factor alone is insufficient to make electromotility as the basis of cochlear amplifier.

VIII. CONCLUSIONS

We confirmed that the oscillatory potential in unstimulated OHCs induced by the extracellular potential can be greater than the receptor potential and that this effect can increase the efficiency of electromotility, as proposed by Dallos and Evans.[15] However, we do not find that this effect can extend the frequency limit of about 10 kHz, which was previously estimated.[7] It turned out that the introduction of an extracellular resistance, which is needed to produce the CM, reduces the RP. Since the significance of prestin and electromotility has been established,[4] our result indicates that there must be another factor, such as fast voltage-gated *K* currents, in the basal turn where the characteristic frequencies exceed 10 kHz.

ACKNOWLEDGMENTS

The authors thank Richard Chadwick and two anonymous reviewers for valuable comments. This research was supported by the Intramural Research Program of the NIDCD, NIH.

[1]D. C. Mountain and A. E. Hubbard, "Computational analysis of hair cell and auditory nerve processes," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York, 1995), pp. 121–156.
[2]M. C. Liberman, J. Gao, D. Z. He, X. Wu, S. Jia, and J. Zuo, "Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier," Nature (London) **419**, 300–304 (2002).
[3]A. N. Lukashkin, M. N. Walling, and I. J. Russell, "Power amplification in the mammalian cochlea," Curr. Biol. **17**, 1340–1344 (2007).
[4]P. Dallos, X. Wu, M. A. Cheatham, J. Gao, J. Zheng, C. T. Anderson, S. Jia, X. Wang, W. H. Y. Cheng, S. Sengupta, D. Z. Z. He, and J. Zuo, "Prestin-based outer hair cell motility is necessary for mammalian cochlear amplification," Neuron **58**, 333–339 (2008).
[5]G. D. Housley and J. F. Ashmore, "Ionic currents of outer hair cells isolated from the guinea-pig cochlea," J. Physiol. (London) **448**, 73–98 (1992).
[6]J. Santos-Sacchi, "On the frequency limit and phase of outer hair cell motility: Effects of the membrane filter," J. Neurosci. **12**, 1906–1916 (1992).
[7]M. Ospeck, X.-X. Dong, and K. H. Iwasa, "Limiting frequency of the cochlear amplifier based on electromotility of outer hair cells," Biophys. J. **84**, 739–749 (2003).
[8]T. Gold, "Hearing. II. The physical basis of the action of the cochlea," Proc. R. Soc. London, Ser. B **135**, 492–498 (1948).
[9]J. Allen, "Cochlear micromechanics—A physical model of transduction," J. Acoust. Soc. Am. **68**, 1660–1670 (1980).
[10]A. A. Spector, W. E. Brownell, and A. S. Popel, "Effect of outer hair cell piezoelectricity on high-frequency receptor potentials," J. Acoust. Soc. Am. **113**, 453–461 (2003).
[11]V. Rybalchenko and J. Santos-Sacchi, "Cl^- flux through a non-selective, stretch-sensitive conductance influences the outer hair cell motor of the guinea-pig," J. Physiol. (London) **547**, 873–891 (2003).
[12]J. Santos-Sacchi, L. Song, J. Zheng, and A. L. Nuttall, "Control of mammalian cochlear amplification by chloride anions," J. Neurosci. **26**, 3992–3998 (2006).
[13]M. Ospeck, X.-X. Dong, J. Fang, and K. H. Iwasa, "Electromotility in outer hair cells: a supporting role for fast potassium conductance," ORL **68**, 373–377 (2006).
[14]P. Dallos and B. Evans, "High-frequency motility of outer hair cells and the cochlear amplifier," Science **267**, 2006–2009 (1995).
[15]P. Dallos and B. N. Evans, "High-frequency outer hair cell motility: corrections and addendum," Science **268**, 1420–1421 (1995).
[16]T. Ren, "Longitudinal pattern of basilar membrane vibration in the sensitive cochlea," Proc. Natl. Acad. Sci. U.S.A. **99**, 17101–17106 (2002).
[17]A. Fridberger, J. B. de Monvel, J. Zheng, N. Hu, Y. Zou, T. Ren, and A. Nuttall, "Organ of corti potentials and the motion of the basilar membrane," J. Neurosci. **24**, 10057–10063 (2004).
[18]P. Dallos and M. A. Cheatham, "Production of cochlear potentials by inner and outer hair cells," J. Acoust. Soc. Am. **60**, 510–512 (1976).
[19]W. S. Rhode, "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231 (1971).
[20]L. Robles and M. A. Ruggero, "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352 (2001).
[21]C. D. Geisler, G. K. Yates, R. B. Patuzzi, and B. M. Johnstone, "Saturation of outer hair cell receptor currents causes two-tone suppression," Hear. Res. **44**, 241–256 (1990).
[22]A. R. Cody and I. J. Russell, "The response of hair cells in the basal turn of the guinea-pig cochlea to tones," J. Physiol. (London) **383**, 551–569 (1987).
[23]The Nernst potential with 140 mM K^+ inside and 2.7 mM K^+ outside the cell.
[24]A. N. Salt, N. Inamura, R. Thalmann, and A. Vora, "Calcium gradients in inner ear endolymph," Am. J. Otol. **10**, 371–375 (1989).
[25]K. H. Iwasa and M. Adachi, "Force generation in the outer hair cell of the cochlea," Biophys. J. **73**, 546–555 (1997).
[26]S. Ramamoorthy, N. V. Deo, and K. Grosh, "A mechano-electro-acoustical model for the cochlea: response to acoustic stimuli," J. Acoust. Soc. Am. **121**, 2758–2773 (2007).
[27]F. Mammano and J. F. Ashmore, "Differential expression of outer hair cell potassium currents in the isolated cochlea of the guinea-pig," J. Physiol. (London) **496**, 639–646 (1996).
[28]C. J. Kros, A. Rüsch, and G. P. Richardson, "Mechano-electrical transducer currents in the hair cells of the cultured mouse cochlea," Proc. R. Soc. London, Ser. B **249**, 185–193 (1992).
[29]C. J. Kros, "Physiology of mammalian cochlear hair cells," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York, 1996), pp. 318–385.
[30]D. Z. Z. He, S. Jia, and P. Dallos, "Mechanoelectrical transduction of adult outer hair cells studied in a gerbil hemicochlea," Nature (London) **429**, 766–770 (2004).

Measuring distortion product otoacoustic emissions using continuously sweeping primaries

Glenis R. Long[b)]
Speech and Hearing Sciences, Graduate Center of the City University of New York, 365 Fifth Avenue, New York 10016 and
Department of Otolaryngology, NYU Langone Medical Center, 550 First Avenue, New York, NY 10016

Carrick L. Talmadge
National Center for Physical Acoustics, University of Mississippi, Mississippi 38677

Jungmee Lee
Speech Language and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 18 December 2007; revised 28 May 2008; accepted 29 May 2008)

Distortion product otoacoustic emission (DPOAE) level from normal hearing individuals can vary by as much as 30 dB with small frequency changes (a phenomenon known as DPOAE fine structure). This fine structure is hypothesized to stem from the interaction of components from two different regions of the cochlea (the nonlinear generator region and the reflection component from the DP region). An efficient procedure to separate these two components would improve the clinical and research utility of DPOAE by permitting separate evaluation of different cochlea regions. In this paper, two procedures for evaluating DPOAE fine structure are compared: DPOAE generated by fixed-frequency primaries versus continuously sweeping primaries. The sweep DPOAE data are analyzed with a least squares fit filter. Sweep rates of greater than 8 s per octave permit rapid evaluation of the cochlear fine structure. A higher sweep rate of 2 s per octave provided DPOAE without fine structure. Under these conditions, the longer latency reflection component falls outside the range of the filter. Consequently, DPOAE obtained with sweeping tones can be used either to get more rapid estimates of DPOAE fine structure or to obtain estimates of DPOAE from the generator region uncontaminated by energy from the reflection region.
 [DOI: 10.1121/1.2949505]

PACS number(s): 43.64.Jb [BLM] Pages: 1613–1626

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) are sounds generated by the healthy inner ear and detected by placing a sensitive probe in the outer ear canal. They are a consequence of the nonlinear properties of the healthy inner ear (amplification of low-level tones but not high-level signals). Two external tones (called primaries) at frequencies f_1 and f_2 (where f_2 is greater than f_1) are inputted through two transducers placed in the ear canal. Distortion components are generated at predictable frequencies not present in the original signal (e.g., $2f_1-f_2$), which can be separated from the input stimuli using spectral analysis.

Many investigators assume that the level of the DPOAE reflects cochlear function near the f_2 place, where there is maximal overlap of the two frequencies on the basilar membrane. However, other investigators now accept that the energy can come from two major regions. The DP is generated near the f_2 region and some energy travels from this place toward the oval window, through the middle ear into the ear canal. This component is often referred to as the "generator" or "overlap" component. When the DP frequency is lower than the f_2 frequency, a proportion of the DP energy, (referred to as the reflection component) behaves like an external tone of the same frequency in that it travels to the characteristic place for that DP frequency and is reflected from there (Shaffer *et al.*, 2003). Because energy traveling from base to the characteristic place is amplified, the component from the DP can be larger than the component from the overlap region (Talmadge *et al.*, 1999). Both of these components are expected to change in phase with frequency, but the rate of phase change is very different for the two components (slow for the generator component and rapid for the reflection component). This means that the vector sum of the two components at the oval window and ear canal can be either greater or smaller than the initially generated component coming from the f_2 region. When they are in phase, the magnitude is greater than the magnitude of the component from the generator region alone; when they are out of phase, the magnitude of the summed components is less than the magnitude of the generator component. The interaction between the two components gives rise to the level fluctuations characteristic of DPOAE fine structure. Whereas the generator component arises from the nonlinear interaction of the activity pattern of two relatively high-level external stimuli, the reflection component comes from the DP place and results

[0)] Parts of this paper were presented a the 27 Midwinter Research Meeting of the Association for Research in Otolaryngology Abs. 1076 2004.
[b)] Author to whom correspondence should be addressed. Electronic mail: glong@gc.cuny.edu

from relatively low-level stimulation at the DP place. Consequently, the reflection component is expected to be much more vulnerable to cochlear insults than the generator component. The presence of DPOAE fine structure is therefore an indicator of cochlear health and can be used for differential diagnosis of outer hair cell damage (Mauermann *et al.*, 1999).

When DPOAE measurements are made at widely spaced frequencies, the investigator does not know where the DPOAE falls within the fine structure. This makes it difficult to determine the health of a single place on the basilar membrane (any DPOAE is the sum of energy from two regions of the basilar membrane). Tools have been developed for removing and/or evaluating the component from the generator region independently from the component from the DP region. One such procedure places suppressor tones near the DP frequency with the intent to suppress the component from the DP place while leaving the generator component unchanged (Heitmann *et al.*, 1996; Johnson *et al.*, 2007; Johnson *et al.*, 2006; Konrad-Martin *et al.*, 2002; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004; Plinkert *et al.*, 1997). Unless one has some idea of the relative amplitudes of the two components, this may sometimes increase the fine structure instead of reducing it (Talmadge *et al.*, 1999).

If one has detailed measures of the amplitude and phase of DPOAE as a function of frequency, one can use estimates of the difference in DPOAE latency based on the group delay of the two components (the component from the overlap region has a shorter group delay than the component from the DP place) to separate the components.[1] An inverse fast Fourier transform (IFFT) of the amplitude and phase data provides an estimate of level as a function of time, leading to two major peaks in the time domain. Filtering in the time domain and returning the stimuli to the frequency domain permit separation of the two components (Shera and Zweig, 1993; Stover *et al.*, 1996). A reliable estimate of the distribution in the time domain depends on obtaining very detailed estimates of the cochlear fine structure (usually a time consuming procedure). The success of this procedure depends on the window and the choice of filter used in the time-windowing procedure used to separate the components (Withnell *et al.*, 2003). If one component is much larger than the other, then temporal splatter can mean that the estimates of the two components may not be completely separated.

This paper describes an attempt to develop a more efficient tool for measuring DPOAE fine structure by using continuously sweeping tones. The two primaries tones are modulated in frequency while maintaining a constant ratio (however, the underlying method can also handle paradigms where the ratios are varied with frequency). If the sweep rate is slow enough, this procedure gives detailed DPOAE measurements that are almost identical to estimates obtained using discrete DPOAE measurements. If the length of the time window is held fixed, more rapid sweeps extract the generator component from the total DPOAE. The ability to separate these two components provides additional support for the two-source theory of DPOAE generation.

The first experiment describes an attempt to develop a more efficient tool for measuring the DPOAE fine structure. The second attempts to test the discovery in the first experiment, that we are able to use the sweeping tone procedure to isolate the generator component.

II. EXPERIMENT 1. COMPARISON OF TRADITIONAL FIXED-FREQUENCY DPOAE WITH DPOAE OBTAINED WITH CONTINUOUSLY SWEEPING PRIMARIES

In order to evaluate whether DPOAE obtained using continuously sweeping primaries are comparable with DPOAE obtained with traditional DPOAE measurements, DPOAE were obtained using both procedures. The traditional measurements were made with the frequency of the primaries fixed within each presentation of the stimuli, and the frequency of the primaries changed for the next stimulus. The frequencies of the primaries were swept within a stimulus presentation for the continuously sweeping paradigm. The frequency ratio and primary levels were kept constant in both procedures

A. Methods

1. Participants

Six members of Ph.D. program at the Graduate Center of the City University of New York (one male, L5, and five females). Subjects were seated in a reclining chair in a double walled IAC booth. Subjects were screened to ensure that they had thresholds of 20 dB HL or better up to 4 kHz. Subjects L2, L3, and L2 have no measurable spontaneous otoacoustic emissions (SOAE), L4 has one near 700 Hz at −9 dB SPL, L6 has four SOAE emissions clearly above the noise floor near 970 (1 dB), 1570 Hz (−5 dB), 1750 Hz (−7 dB) and 1950 Hz (2 dB), and L1 has several emissions clearly above the noise floor approximately at 1065 Hz (−1 dB), 1395 (−14 dB), 1465 (−11 dB), 2020 (10 dB), and 2455 Hz (−6 dB).

2. Stimuli

The stimuli had a fixed $f_2/f_1=1.22$ and $L_1=65$ dB SPL and $L_2=55$ dB SPL in all instances. Stimuli were calibrated using Kemar to get an estimate of the level at the eardrum in order to avoid problems generated by calibration at the entrance to the ear canal (Siegel, 2007). At the beginning, middle, and end of each session, white noise was played through each tube phone in turn, recorded and analyzed using a FFT in order to evaluate the probe fit and ensure that levels near 1000 Hz approximate the required stimulus levels.

(a) Fixed-frequency DPOAE. The procedure for collecting fixed-frequency DPOAE was similar to that reported in Talmadge *et al.* (1999). The stimuli were 4 s tones, with a 5 s pause every five stimuli (to give the subjects an opportunity to swallow, clear their throats, and do any noisy activities). Stimuli were presented to generate DPOAE from 642 to 2568 Hz in 6 Hz steps (f_2 from 1002 to 4020 Hz). The stimuli were divided into three separate files of approximately 108 stimuli each and the order of the stimuli pairs

($f_2/f_1=1.22$) within each file was randomized. The total data collection time was about 35 min.

(b) Continuously Sweeping DPOAE. Three overlapping one-octave linear-frequency sweeps were used to cover a two-octave range. F_2 was swept from 1000–2000, 1414–2828, and 2000–4000 Hz in 2, 8, or 16 s. Upward and downward sweeping primaries were used to determine whether the direction of the sweep influenced the results. The three overlapping upward sweeps were played with increasing starting f_2 frequency, followed by the three downward sweeps (from high to low). A 5 s break was introduced after, nine 2 s sweeps, three 8 s sweeps, or one 16 s sweep, so that the subjects had a 5 s break every 16–24 s. The number of sweeps obtained in each session depended on the sweep duration; 108 for the 2 s/octave sweeps (taking 5 min), 84 for the 8 s/octave sweeps (taking 14 min), and 60 for the 16 s/octave sweeps (taking 21 min). Note that the time taken for sweeping tone measurements provides two estimates of the fine structure in comparison with the single estimate for the fixed-frequency data.

3. Procedure

Custom programs for Mac computers (MAC OS X) generated the stimuli and recorded the ear canal signals. Three ER2 tube phones were connected to a three-port ER10A microphone, which was inserted in the ear canal (the third tube phone was needed in Experiment 2). Before being digitized by the Motu 828 (24 bit, 44100 samples/s), the signal from the microphone was conditioned and preamplified by the Etymotic preamplifier (gain=1) and then amplified (gain =100) and filtered (300–10,000 Hz) by a Stanford SR560 low-noise amplifier under computer control.

When possible, the fixed-frequency DPOAE and the continuously sweeping DPOAE with the three different sweep rates were obtained in one session. Fixed-frequency data were obtained on two different days for L4. The middle-frequency range of fixed-frequency data was run on both days to provide an indication of the stability of the measures. The fixed-frequency data and the sweep data from subjects L5 and L6 were obtained on different days (the dates are indicated on the plots).

4. Analysis

(a) Fixed-frequency DPOAE. The least-squares-fit (LSF) procedure described in Long and Talmadge (1997), Talmadge *et al.* (1999) and Dhar *et al.*, (2002, 2005) was used to extract the stimuli. The 4 s sound files were reduced to 1/2 s duration by temporal averaging of overlapping segments (step size of 1/6 s). When one uses a LSF procedure, the magnitude and phase of the expected components are adjusted to minimize the sum of squares of the differences between the fitted wave form and the ear canal sound file. A LSF procedure was used instead of Fourier analyses because of the following reasons:

(1) It is more efficient; it only evaluates the frequencies of interest.
(2) Stimulus frequencies do not need to be limited to frequencies falling in the center of the FFT analysis window.
(3) It provides better separation of components when they are close together (the spectral splatter is less).
(4) It permits better separation of signal and noise when analyzing nontonal signals. The estimate of the DPOAE using a FFT is the sum of the noise and any stimuli falling in the analysis bin. The estimate of the DPOAE using the LSF contains only the energy that is coherent with the stimulus (the noise is not coherent with the signal).

When the stimuli are well above the noise floor and the components are well separated, the results of the LSF procedure are identical with the results from the FFT. The LSF procedure gives better estimates at low signal levels and can even obtain an estimate (with appropriate phase) below the level of the noise floor. Estimates are obtained from each analysis window. We used 1/4 s analysis windows with a step size of 1/25 s. The median of the resulting level and phase estimates obtained were taken as the measure for each stimulus pair (the median was chosen because it is less affected by short duration artifacts which normally increase the level of the noise floor). The noise floor was obtained in the same way except that alternating temporal windows were inverted to cancel any signal related components. Points with the noise floor above −10 dB SPL were eliminated.

(b) Continuously sweeping DPOAE. Sweeps with high noise levels were eliminated as long as a minimum of eight sweeps could be averaged. The remaining sweeps with identical stimulus conditions (direction and duration) were averaged to reduce the noise floor and subtracted to estimate the noise floor. Up- and downsweeps were analyzed independently to ensure that the direction of the sweep did not affect the data and to provide an indication of the reliability of the procedure.

The LSF analysis described in Talmadge *et al.* (1999) was modified to include both linearly and logarithmically swept tones, and used to extract the levels of the primaries and DPOAE. Overlapping Hann-windowed segments of data were analyzed using the LSF procedure (which models both the swept stimuli and the DPOAE components) to obtain the amplitude and phase of the DPOAE as a function of frequency. The result of the LSF procedure is a bandpass filter that changes center frequency as the DPOAE frequency changes, and produces both in-phase and quadrature components of the signal of interest (this allows both total amplitude and phase to be estimated). The bandwidth of the filter is determined by the size (or duration) of the analysis widow. In this experiment, the analysis window was 1/2 s (22 050 points), which gives a bandwidth of 2 Hz (sampling rate/analysis window). The analysis bandwidth can be increased by reducing the size of the analysis window. When sweeping tones are used, increasing the bandwidth includes components with longer delays inside of the filter window (see discussion).

The step size between analysis windows was typically 1/20 s. This determines the distance between data points. Additional analysis was done with steps of 1/80 s. Changing

the step size did not significantly change the pattern of results. The use of such a long analysis window means that the obtained DPOAEs no longer cover the total stimuli range. The proportion maintained depends on both the analysis window and sweep duration.

The noise floor was obtained by inverting every second sound file, before executing the same analysis to obtain estimates of the noise floor at the frequencies of the DPOAE and primaries. In order to do this, one needs an even number of files in order to cancel any stimulus related components.

(c) Separating DPOAE generator and reflection components. In some cases, it was necessary to separate the generator and reflection components of DOAE. The methodology used here was based on that used by a program called NIPR, written by the second author. NIPR is a MATLAB-based analysis framework that can be used to study the various DPOAE components, using an IFFT-based algorithm described in Dhar *et al.*, (2002) and Withnell *et al.* (2003). Basically, this method uses the IFFT to convert the frequency domain complex-valued DPOAE amplitude to the timedomain, where a time-window filter is applied. In a linear system, this inverse transform corresponds to the system impulse response function (IPR). The time-window filter typically bandpass filters around the generator component of the DPOAE (which is usually near zero lag).

In order to reduce windowing artifacts, multiple overlapping windows are computed and averaged together (hence NIPR signifies an *n*-window IPR). The analyses performed in this paper were based on a C-language predecessor of NIPR. The main functional difference in the algorithms between the two software packages is that the filter for the generator component is preset in the C-version of the code, whereas it is robustly selected in the MATLAB code. The details of the IPR analysis used in this paper will be discussed in greater detail below.

B. Results

The similarity of the DPOAE level and phase estimates obtained with fixed-frequency primaries and those obtained from the slowest continuously sweeping primaries (16 s/octave) can be seen by looking at the data in Fig. 1. The circles indicate the individual fixed-frequency DPOAE, while the lines indicate the continuously sweeping DPOAE. We chose not to use symbols for the sweeping DPOAE because of the finer resolution obtained using the sweeps (1, 2, 4 Hz, respectively, for the 1000–2000, 1414–2828, and 2000–4000 Hz sweeps) and the lack of independence of adjacent points. The DPOAE phases from the same stimuli are displayed in Fig. 2. Data from upsweeping primaries are shown with solid gray lines and data from downsweeping primaries are shown with dashed gray lines. It is mostly not possible to detect the downswept data (dashed gray lines) because the DPOAE obtained with up-and downsweeps tends to overlie each other. Note that the three overlapping sweeps used in the experiment mean that some frequencies are represented twice. The results were so similar that one is mostly unable to tell that the lines overlap. Even though subjects were not selected on the basis of their fine structure,

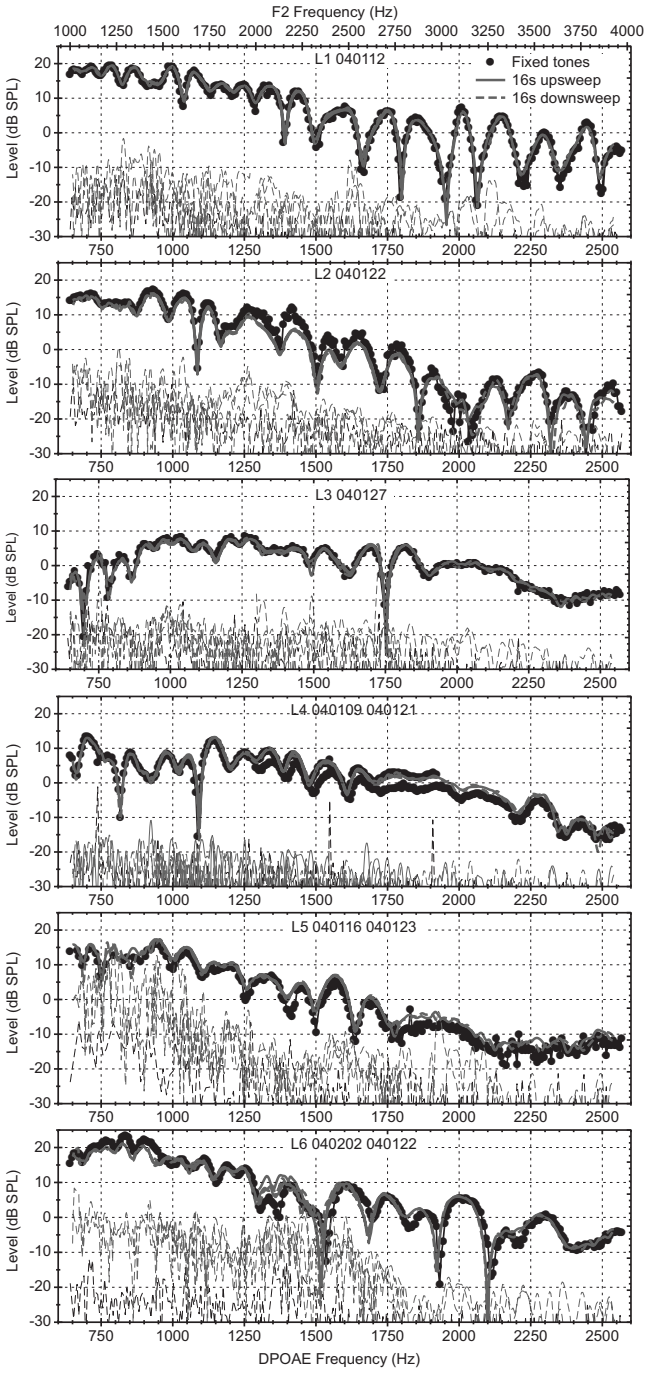


FIG. 1. Estimates of DPOAE level as a function of frequency obtained using fixed-frequency paradigm (filled circles) and 16 s/octave linear sweeps (upsweeps solid gray lines, downsweeps dashed gray lines—this convention will be used in all the following figures). The noise floors for the two different paradigms, are indicated by the thin black and gray lines. The 16 s/octave data are repeated in Figs. 3 and 5.

all of them show distinct fine structure, which is different for each subject. In all cases, the DPOAE levels and phases obtained using the different procedures are nearly identical when the data were collected on the same day. The dates on which the data were collected are indicated in the top right of each panel.

The impact of obtaining measurements on different days can be seen in L4's data (Figs. 1 and 2, fourth panel). The fixed-frequency data were obtained during two different ses-

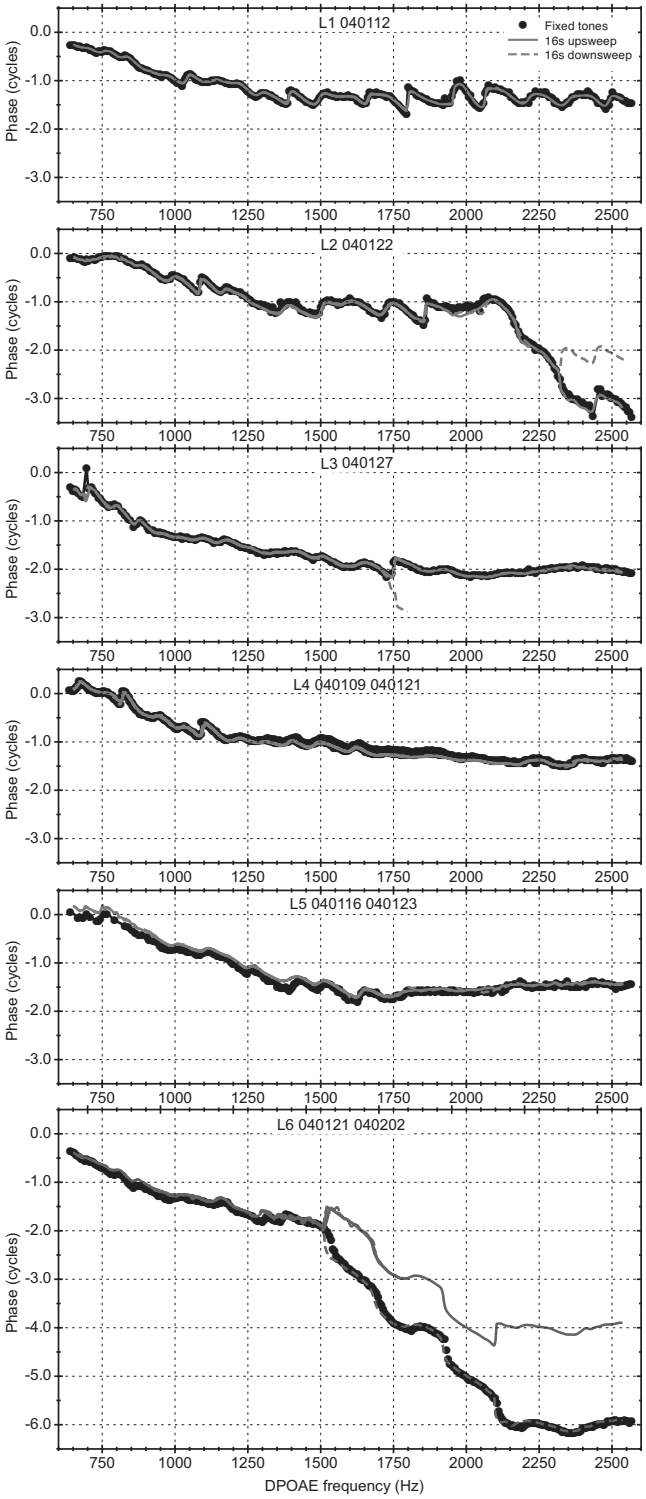


FIG. 2. Estimates of DPOAE phase as a function of frequency obtained using fixed-frequency paradigm (black circles and lines) and 16 s/octave linear sweeps (solid gray lines for upsweeps and dashed gray lines for downsweeps). Note that the scale for the bottom panel is the same as the other panels even though the size is different. The 16 s data are repeated in Figs. 4 and 6.

sions separated by 12 days. The middle-frequency range was repeated on both days and the data from both days are shown in Figs. 1 and 2. The level of the DPOAE in this frequency region (Fig. 1) is somewhat lower on the first session but the pattern of the fine structure is not different. There are much smaller differences in the phase data (Fig. 2) indicating that the phase was much less sensitive to differences in probe position. The fixed-frequency data and sweep data were collected on separate days for subjects L5 and L6 (bottom two panels), in both these cases the pattern of data on the two days is very similar. The probe must have been placed in a similar position.

There are few places where the phase appears to differ between up-and downsweeps (see Fig. 2, panels 2, 3, and 6). The separation of the phase trajectories always occurs when there is a rapid change in phase associated with a deep threshold minimum. This pattern occurs when the two components are nearly equal in level as detailed in Talmadge *et al.*, (1999). A small change in relative amplitude can lead to either a rapid increase in phase (generator component dominant) or a rapid decrease in phase (reflection component dominant). In our data, the difference is always between the two estimates with continuously sweeping DPOAE, and never between sweeping DPOAE and fixed-frequency DPOAE.

Note that the rate of phase change for each subject, and for different frequency ranges within a subject, can be very different. The rate of phase change above 1500 Hz for subject L6 (bottom panel) was much larger than the 3 cycles/octave seen in the other subjects, so that the size of the panel had to be modified in order to cover 6 cycle/octave change seen in this subject (the scale does not change). Subject L2 (second panel) shows steep phase change above 2100 Hz.

Given the similarity of the 16 s/octave data to the fixed-frequency data, we have chosen to compare the 16 s/octave DPOAE with DPOAE obtained using more rapid sweep rates. The 16 s/octave data from Fig. 1 is, therefore, repeated in Figs. 3 and 4. The 16 s/octave data are plotted with a thin light gray line, the 8 s/octave data with a think black line, and the 2 s/octave data with the middle width dark-gray line. Increasing the sweep rate from 16 s/octave to 8 s/octave (effective analysis step sizes, of 4, 6, and 8 Hz for the different frequency regions) had little impact. The DPOAE level (Fig. 3) obtained with the different sweep rates differs near the deepest minima producing either a reduction (L1, L3, and L6, first and third panels of Fig. 3, open arrows) or enhancement of the DPOAE fine structure (L6 bottom panel Fig. 3, filled arrows). The phase changes are mostly small but at times the pattern changes from a steep drop in phase in the 16 s/octave data to an increment in phase in the 8 s/octave data (see open arrows in panel 6 of Fig. 4).

Increasing the sweep rate to 2 s/octave (analysis step sizes of 6, 22, and 32 Hz) had a much more dramatic effect. This can be seen in Figs. 3 and 4 with the middle width dark gray line. We can no longer detect any fine structure in the DPAOE levels (Fig. 3) and the phase (Fig. 4) has a similar shallow slope for all subjects. Note that the range of frequencies obtained is less for the 2 s sweeps than the 16 s sweeps.

C. Discussion

The results show that the slower continuously sweeping tones can be used to get more detailed information about

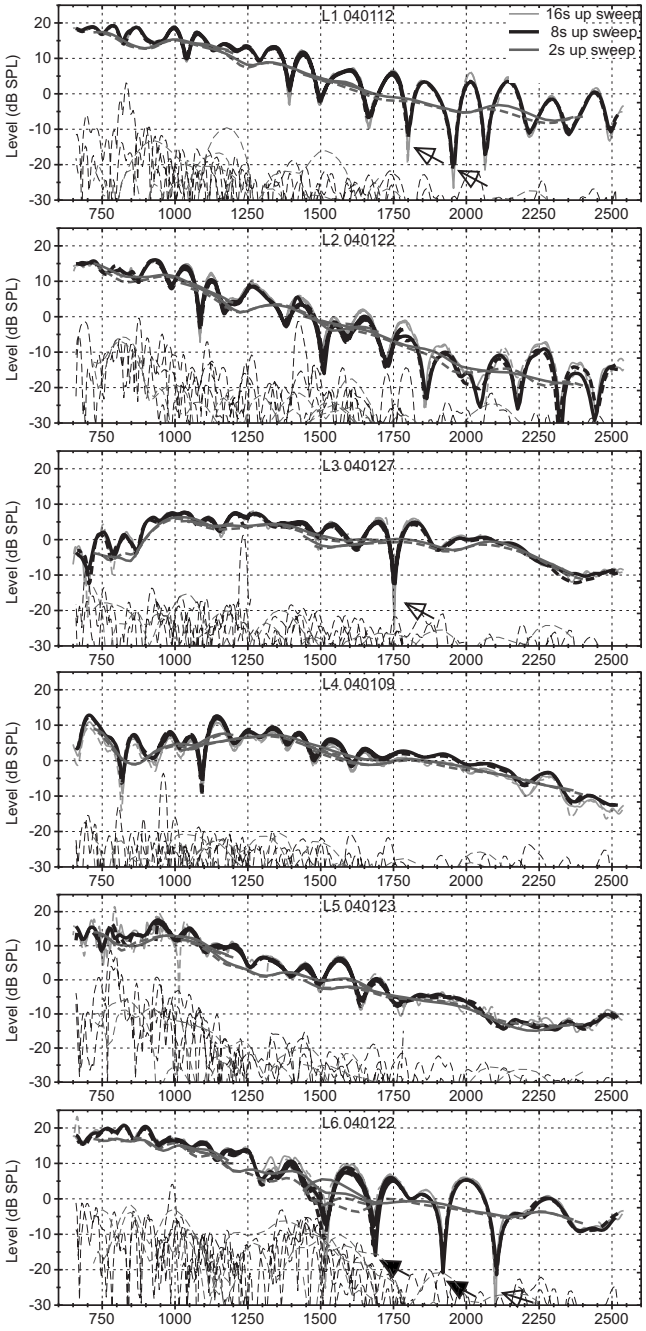


FIG. 3. Estimates of DPOAE level as a function of frequency from the 16 s/octave sweeps (thin, light gray lines), 8 s/octave sweeps (wide, black lines), and 2 s/octave sweeps (middle width, dark gray lines). Data obtained using upsweeps are shown with solid lines and data obtained with downsweeps are shown with dashed lines. The 16 s data from Fig. 1 are repeated in this figure. The arrows are explained in the text.

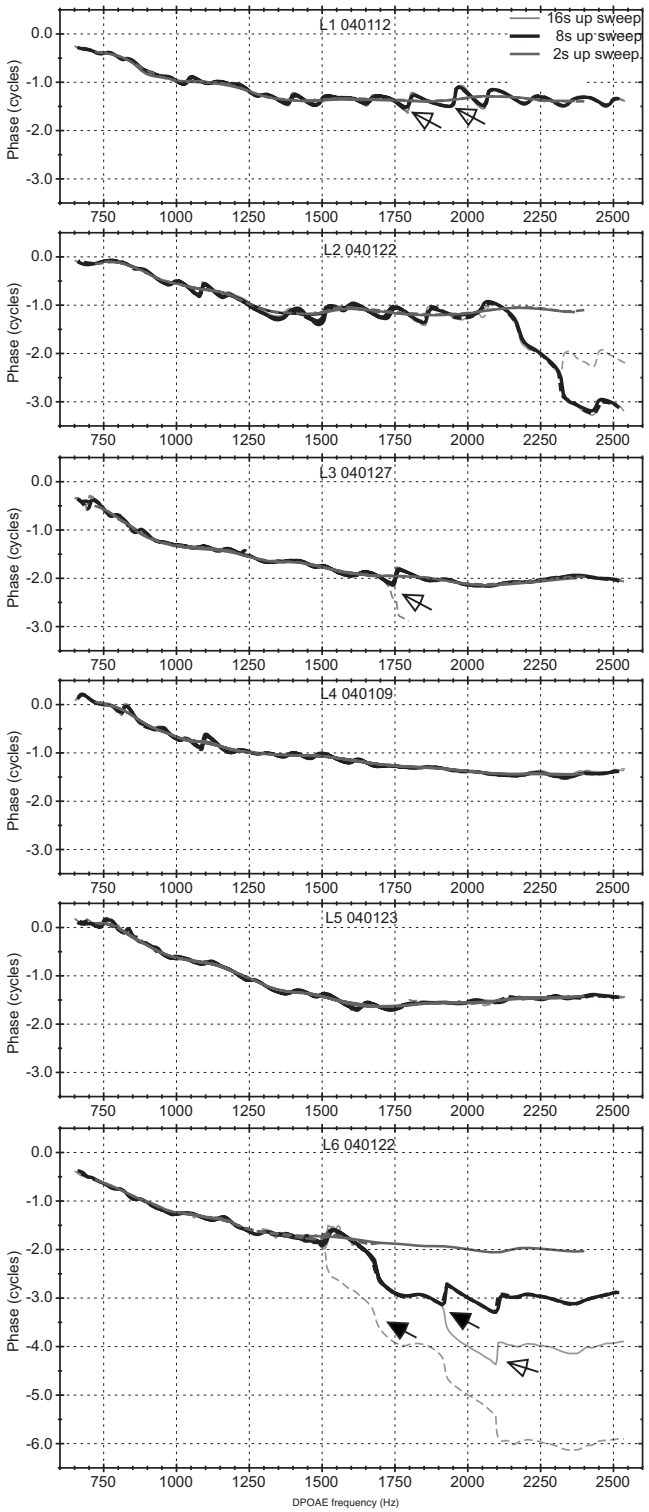


FIG. 4. Estimates of DPOAE phase as a function of frequency from the 16 s/octave sweeps (thin, light gray lines), 8 s/octave sweeps (wide, black lines), and 2 s/octave sweeps (middle width, dark gray lines). Data obtained using upsweeps are shown with solid lines and data obtained with downsweeps with dashed lines. The 16 s data from Fig. 2 are repeated in this figure. The arrows are explained in the text.

DPOAE fine structure in less time. We obtained two estimates of the fine structure in 21 min with the 16 s/octave sweeps while it took 35 min to get one estimate with the fixed-frequency data. The 8 s/octave sweeps gave very similar data in 14 min. The greater resolution we obtained permits more reliable evaluation of the depth of the fine structure and will permit better extraction of the amplitudes of the different components, when this is done by performing an IFFT on the data and filtering in the time domain (IPR).

The pattern seen for the 2 s/octave sweeps was not anticipated, but can be easily interpreted by (and, in fact, is predicted by) the two-source model. The LSF procedure acts as a moving narrow band filter. The center frequency of this filter moves in time, coupled with the frequency changes in the analyzed stimulus. Consequently, any component that comes back at that frequency with a longer delay no longer

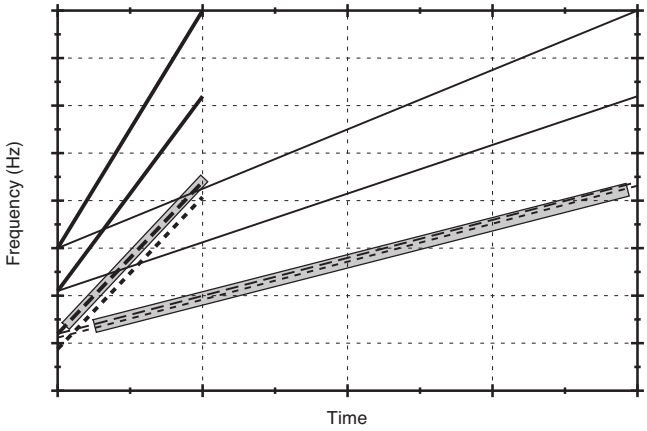


FIG. 5. Schematic representation of the sweeping primaries (solid lines), and undelayed (dashed lines) and delayed (dotted lines) DPOAEs. For a given analysis window length of analysis window which determines the filter bandwidth (*df*) will be the same for both chirp rates (shaded boxes), and the latency difference between the initial source(s) and the reflection source(s), *dt*, at each frequency is independent of the sweep rate. This leads to a separation of the components in the case of fast sweep rates, while the two DPOAE components both fall within the filter for slower rates.

falls within the filter bandwidth, and will not be detected by the analysis. The shorter latency generator component always falls within the filter bandwidth used in this experiment. The rate of stimulus change (and thus filter change) for the 16 s/octave sweeps means that both components fall within the filter (the reflection component returns within the bandwidth of the filter). In contrast, when the frequency is rapidly changing (2 s/octave), the filter (which moves with the predicted DPOAE frequency) will no longer include the longer latency reflection component. Consequently, the only component detected by the analysis will be the generator component. It appears that the fast sweep permits extraction of the generator component in the time domain without the addition of suppressor tones. Figure 5 provides a schematic demonstration of rapid and slower sweeps. The upper panel shows the primaries (solid lines), and undelayed (dashed lines), and delayed (dotted lines) DPOAEs. The delays shown in the figure are the same for the two sweep rates, but are not physiological in that they are longer than DPOAE delays and do not vary with frequency. Note that while it is easy to see the two DPOAE traces for the faster sweep, they almost overlap for the 8 s/octave sweep. The lower panel adds shaded areas representing the analysis filters. The width of the filter is the same for two sweep rates. While both components fall within the filter for slower sweep, the filter only covers the undelayed component of the rapid sweep. (Note that by increasing the filter bandwidth, or equivalently reducing the duration of the filter, both components that will fall within the filter can still be detected by the same swept-frequency filter, at the cost of elevation of the noise floor. See below for more evidence.)

The apparently unpredictability of the difference in magnitude of DPOAE fine structure as the sweep rate increased from 16 to 8 s/octave (open versus filled arrows in Fig. 3) can be explained by this hypothesis. While this sweep rate is not fast enough to move the filter so that it entirely misses the reflection component, the filter may reduce the apparent magnitude of the reflection component. This reduction of the magnitude of the reflection component will have the greatest effect on the fine structure when the levels of the components are nearly equal (when the fine structure minima are deepest—see arrows in Fig. 3). A small change in the amplitude of the reflection component under these circumstances could lead to either a reduction of the depth of the minima (when the generator component is dominant, see open arrows in Fig. 3) or a deeper minimum (when the reflection component is dominant, see filled arrows in Fig. 3). The pattern of phase change with frequency can provide a clue as to which component is dominant and this can predict the change in fine structure with suppression of the DPOAE by a third tone (Dhar and Shaffer, 2004; Talmadge *et al.*, 1999). This approach can also be used to interpret the change in fine structure as the sweep rate is increased from 16 to 8 s/octave. The 8 s/octave sweeps generate deeper fine structure in L6 (Fig. 4, bottom panel filled arrows) in the frequency region with steep phase slopes, and shallower minimum in subjects such as L1 and L6 where the phase increases instead of decreasing (Fig. 4, open arrows). This is consistent with predictions from the two-source model and thus provides support for the two-source model.

The best way to test the hypothesis, that the DPOAE obtained using the 2 s/octave sweeps provide an estimate of the generator component uncontaminated by the reflection component, would be to compare the DPOAE level obtained using this procedure with those obtained using the two other procedures described in the literature: suppression by a third tone near the DP frequency and filtering in the time domain. While suppression by a third demands collection of additional data (which is done in Experiment 2, filtering in the time domain can be done on the data from this experiment.

A comparison of DPOAE estimates using the 2 s/octave sweeps and estimates of the generator component extracted using the IPR analysis on the fixed-tone data from this experiment can be seen in Figs. 6 and 7. The fixed-frequency data were analyzed using the procedure described in detail in Dhar *et al.* (2002) and Withnell *et al.* (2003). The amplitude and phase data from the fixed-frequency data were windowed using overlapping Hann windows of 400 Hz (100 Hz steps). Each window was converted into the time domain using an IFFT. Examples of time domain data obtained in this way when analyzing data from Experiment 2 can be seen in the left column of Fig. 8 (thin line). Two major peaks could usually be seen—one with a very short latency (solid line) and one with a longer latency (dashed line). The early peak is assumed to stem primarily from the generator component and the later peak from the reflection component. Once the data were in the time domain, it was bandpass filtered (−3 to 4 ms) to extract the generator component. The filtered data were then converted back into the frequency domain and the data from the different windows recombined to provide estimates of the generator component alone over the complete frequency range. Note that the pattern of DPOAE level obtained from the 2 s/octave sweeps is very similar to the generator component estimates obtained from the fixed-frequency data (see Fig. 6). The greatest differences are seen when the fixed-frequency data were obtained on a

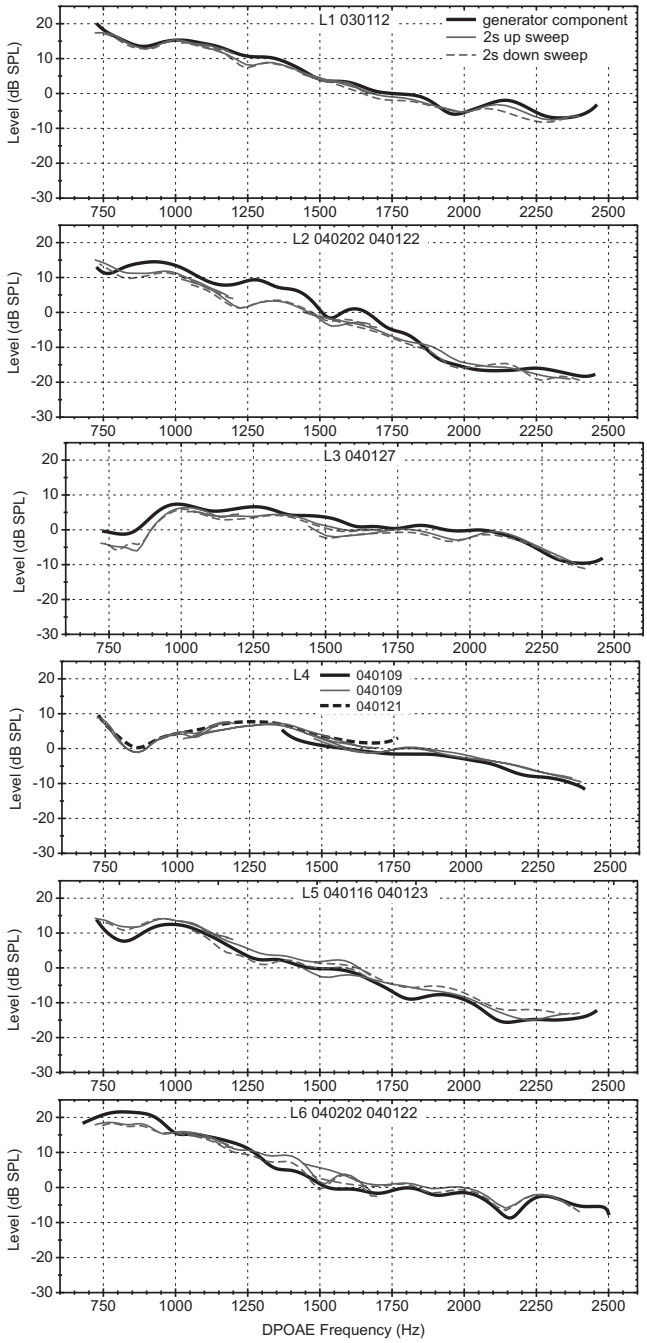


FIG. 6. Estimates of the level of the generator component level obtained from time domain filtering of the IFFT of the amplitude and phase of the pure-tone DPOAE data (black lines) compared with the DPOAE level obtained from the 2 s/octave sweeps (gray lines). Data obtained using upsweeps are shown with solid lines and data obtained with downsweeps with dashed lines.

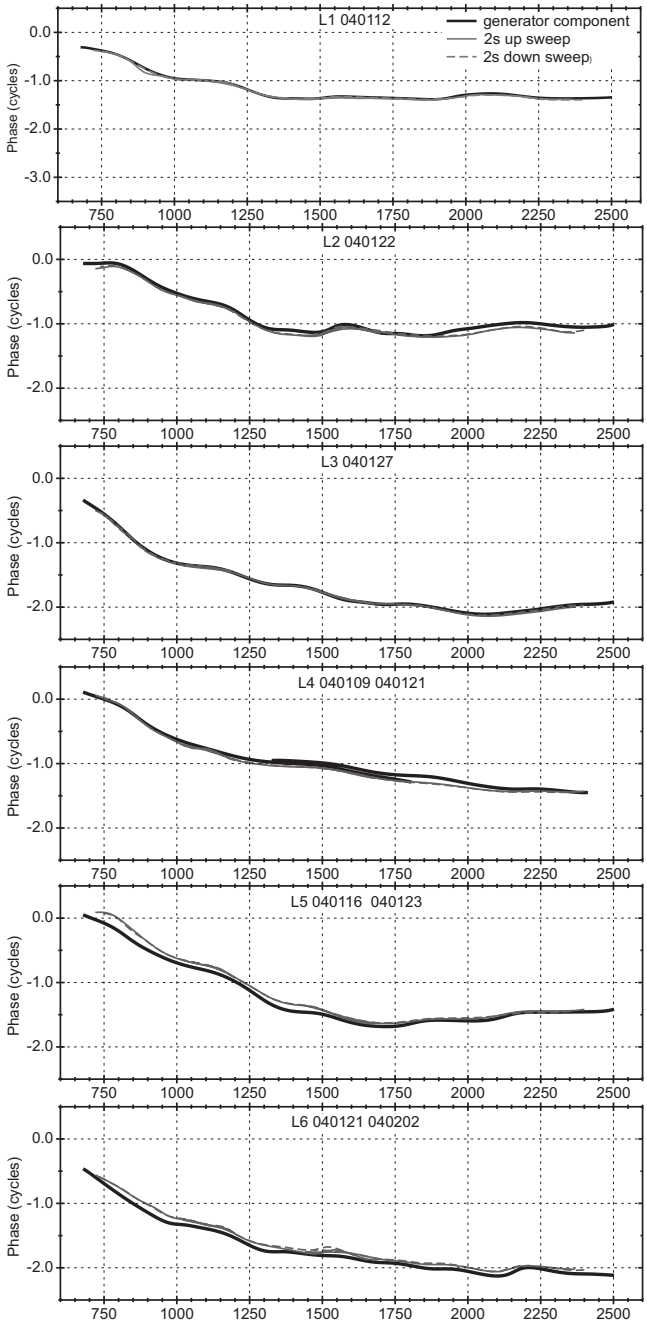


FIG. 7. Estimates of the generator component phase as a function of frequency obtained in the same manner as the estimate of generator component level in Fig. 6 (black lines) compared with the DPOAE phase obtained from the 2 s/octave sweeps (gray lines). Data obtained using upsweeps are shown with solid lines and data obtained with downsweeps with dashed lines.

different day than the continuously sweeping tone data (subjects L2, L4, L5, and L6). Differences in the probe fit can lead to differences in the stimulus at the ear drum and thus differences in DPOAE level. Other differences between the two estimates could stem from temporal splatter in the time domain preventing complete separation of the two components. Note that the phase differences (Fig. 7) between the generator component estimates from the two paradigms are much smaller that the amplitude differences; phase is less influenced by differences in probe position.

The similarities between the data obtained from the 2 s/octave sweeps and the extraction of the generator component from the fixed-frequency data provide support for the claim that the 2 s/octave continuous sweep provides an efficient procedure for extracting the generator component across a wide range of frequencies. Additional support for this claim can be obtained by comparing DPOAE estimates obtained using the 2 s sweeps with fixed-frequency DPOAE obtained with the addition of a third tone near the DP frequency to suppress the reflection component (Dhar and Shaffer, 2004; Heitmann *et al.*, 1996; Johnson *et al.*, 2006; Johnson *et al.*, 2007; Konrad-Martin *et al.*, 2002; Konrad-

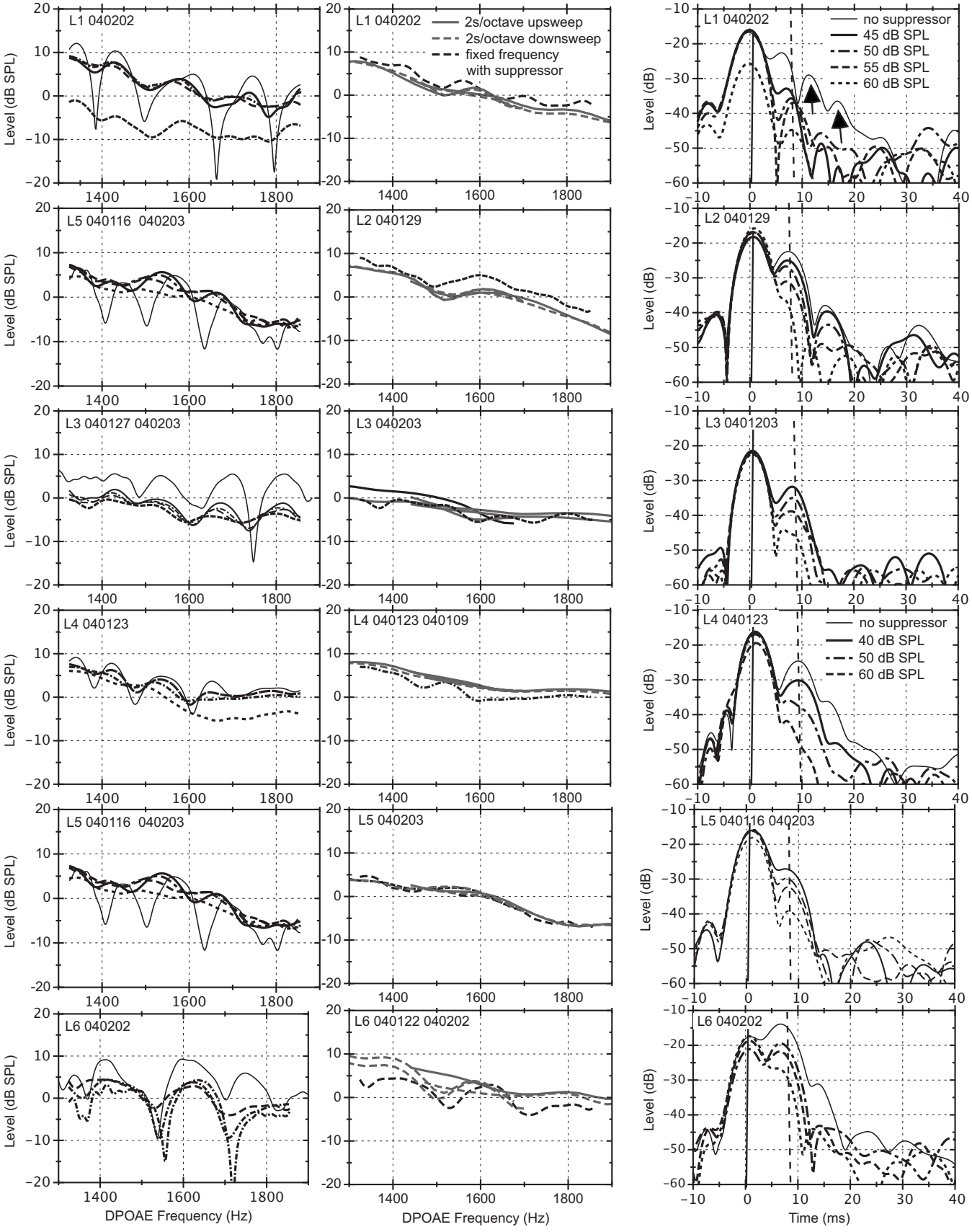


FIG. 8. Left column: DPOAE levels obtained using the fixed-frequency paradigm with suppressor tones 25 Hz above the DPOAE frequency (see legend for suppressor tone levels—a different legend is given for L4 because different suppressors were used for this subject). Center column: The suppressor level in the left hand panel which provided optimal suppression of the reflection component is repeated with the DPOAE level estimates obtained with 2 s/octave sweeps (upsweeps are shown with solid lines and data obtained with downsweeps with dashed lines). Right column: Time domain estimates of DPOAE level obtained from IFFT analysis of a 400 Hz segment of the amplitude data shown in the left panel of this figure and the phase data in Fig. 9. Suppressor tone levels are indicated by the legends in the left hand column.

Martin *et al.*, 2001; Mauermann and Kollmeier, 2004; Plinkert *et al.*, 1997). Similarities between the suppressed data and the 2 s/octave sweeps will provide an additional test of the hypotheses that the 2 s/octave octave sweeps provide an estimate of the generator component as a function of frequency.

III. EXPERIMENT 2. COMPARISON OF ESTIMATES OF THE GENERATOR COMPONENT OBTAINED USING THE CONTINUOUSLY SWEEPING PRIMARIES WITH ESTIMATES OF THE GENERATOR COMPONENT OBTAINED BY SUPPRESSION

A. Methods

1. Participants

The participants were the same as experiment 1.

2. Stimuli

The stimuli were the same as those used to obtain fixed-frequency DPOAE in the first experiment except that a third tone was added at a frequency 25 Hz above the DP frequency. This tone was generated in a third channel and played though a separate ER2 into the three-port ER 10. Suppressor levels were 45, 50, 55, and 60 dB SPL for all subjects except L4 for whom the suppressor levels were 40–60 dB SPL in 10 dB steps. In order to complete data collection in one session, only the middle-frequency region (f_2=2000–3000 Hz) was evaluated. This frequency region was chosen based on the data in experiment 1. This region increases the probability of detecting fine structure while optimizing signal to noise ratio. Due to computer problems, it was not possible to collect unsuppressed data from L3 and L5 on the same day as the suppressed data. Consequently, the pure tone data from experiment 1 were used to provide a basis of comparison. Where possible, the 2 s/octave sweeping DPOAE (without a suppressor tone) was obtained on the same day as the three-tone data (probe slippage prevented this in L6 and the data from experiment 1 were used instead).

3. Analysis

The analysis was the same as experiment 1 except that the amplitude and phase of the suppressor tone and any possible DPOAE generated by the interaction of that tone with other sounds (tones or DPOAE) were added to the LSF analysis of the fixed-frequency data to ensure that they did not contaminate the estimates of the DPOAE. Fluctuations in the fixed-frequency data stemming from signal to noise problems and probe slippage were smoothed by using the NIPR analysis and bandpass filtering in the time domain (−10 to 40 ms). This procedure is described in detail by (Dhar *et al.*, 2002).

B. Results

The impact of the addition of the third suppressor tone on the DPOAE level and phase of the fixed-frequency DPOAE can be seen in Figs. 8 and 9, respectively. The data have been split into two columns (left and center in Fig. 8 and left and right in Fig. 9) to prevent the individual plots from becoming too cluttered. The left column shows the DPOAE (level in Fig. 8 and phase in Fig. 9) as a function of frequency when no suppressor was present (thin line) and when a third suppressor tone was added (different suppressor levels are indicated by different line types). The comparison of the suppressed fixed-frequency DPOAE and the 2 s/octave sweeping DPOAE are shown in the central panel of Fig. 8 and left panel of Fig. 9. The right hand column of Fig. 8, shows plots of DPOAE level as a function of time obtained from the IFFT of the data in the right hand columns of Figs. 8 and 9 (the frequencies analyzed to generate these plots covered 1390–1790 Hz for the majority of subjects, but 1490–1890 Hz for L2). The different line types show the level of the suppressor tone (the legend is in the left hand column).

When the unsuppressed and suppressed data were obtained on the same day, the depth of the DPOAE fine structure was initially reduced without a reduction of the overall level of the DPOAE. Increases in the level of the suppressor tone further reduced the fine structure. In some subjects, L1 and L4, the highest suppressor level led to a reduction of the overall DPOAE level (indicating that the generator component was being suppressed). In contrast, subject L6 still had significant fine structure at this suppressor level (indicating that the reflection component was not fully suppressed). In the two subjects (L3 and L5) on which the unsuppressed data were obtained on separate days, the same pattern of change in fine-structure depth can be seen, but the unsuppressed data differ in overall level from the suppressed data. This probably stems from differences in the probe fit on the different days seen in experiment 1.

The center column in Fig. 8 shows the DPOAE from the 2 s/octave sweeps and the DPOAE from the right hand panel obtained with the highest suppressor level that did not lead to an overall reduction of the overall DPOAE level. We chose these data to optimize the suppression of the reflection component while minimizing any suppression of the generator component. Note the similarity of the levels obtained from the two procedures. In most cases, the sweep DPOAEs did not differ from that of the suppressed fixed-frequency data by more than the measure-to-measure variability of the sweeps alone.

The phase of the DPOAE plotted in the first two columns of Fig. 8 are plotted in Fig. 9. As in experiment 1, the rate of phase change with frequency in the absence of the suppressor and at low suppressor levels for subject L6 (bottom panel) was very large. The phase flattens as the suppressor levels are increased, and at the higher suppressor levels the pattern is one of the fluctuations about a background of slowly changing phase. The other subjects start with this pattern with the depth of the fluctuations decreasing as the suppressor level increases. The left hand panel of Fig. 9 shows the phase changes with frequency associated with the level data in the center panel of Fig. 8, providing a comparison of the phase obtained from the 2 s/octave sweeps with the suppressed fixed frequencies. Note the similarity of the phase trajectories obtained with the two procedures. The continued phase fluctuations in the fixed-frequency data (black lines) in L1 and L6 (the top and bottom panels) reflect

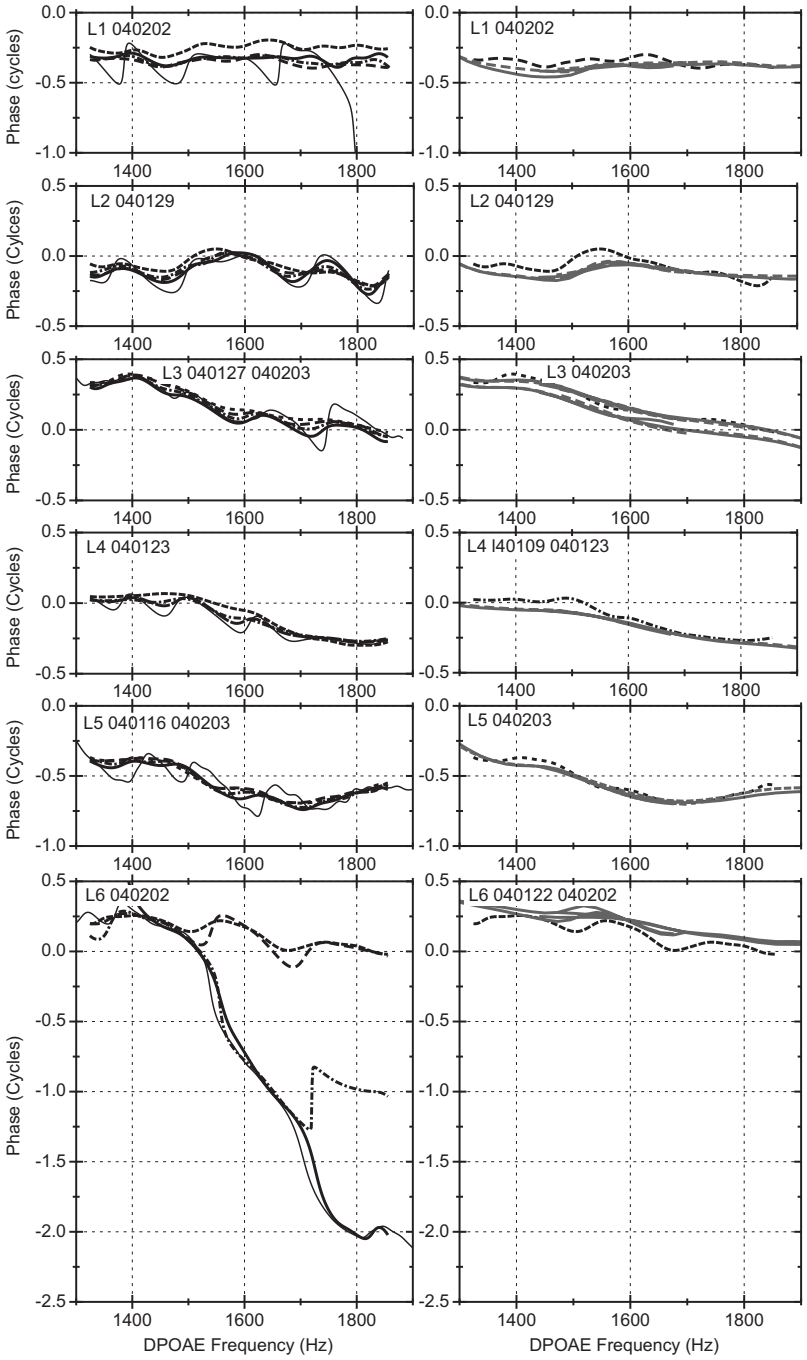


FIG. 9. DPOAE phases associated with the DPOAE level data in Fig. 8. Left hand panel: DPOAE phase obtained using fixed-frequency paradigm with suppressor tones between 45 and 60 dB SPL (see legend in Fig. 8). Right column: The comparison of the suppressed fixed-frequency data with the phase estimates from the analysis of the 2 s/octave sweeps.

the failure to completely suppress the reflection component at the suppressor levels presented for these subjects.

The IFFTs seen in Fig. 8 (left panel) help interpret the impact of the suppressor on the two components. In the top panel (L1), we see that the early (generator) component (near the solid line) remains stable until the highest-level suppressor is added. This level of suppression reduced the magnitude of this peak. When there is no suppressor, there are multiple later peaks in this subject (indicated by arrows) that can be associated with multiple internal reflections. Once a suppressor is added, the later components are no longer seen (Talmadge *et al.*, 1999) and the pattern is reduced to the two peaks predicted by the two-source theory. The second (reflection) component reduces in level as the suppressor level is increased. The pattern is simpler for subjects L2, L3, L4, and L5 in the next four panels. The IFFTs in this frequency region are dominated by two major peaks. The first is relatively unaffected by the primary and the second reduces in level as the suppressor increases in level. In most cases, it is not completely suppressed at the highest suppressor level used. In the subject with the steep phase gradient (L6 bottom panel), the second component is larger than the first component in the unsuppressed data. The addition of the suppressor reduces the level of the second component so that it becomes smaller than the first, but is always larger than the levels for the equivalent conditions in the other subjects.

C. Discussion

DPOAE level and phase obtained using the 2 s/octave sweeps and fixed-frequency stimuli with the addition of a

third suppressor tone are very similar. This helps support the claim that the 2 s/octave sweeps provide an efficient procedure for estimating the generator component uncontaminated by the reflection component. Using a third tone to suppress the generator component is less predictable than the 2 s/octave sweep. The same suppressor level reduced the level of the generator component in one subject (L1) and failed to suppress the reflection component in another subject (L6), where the reflection component was larger than the generator component. Similar intersubject variability has been reported by other investigators (Dhar and Shaffer, 2004; Johnson *et al.*, 2007; Johnson *et al.*, 2006; Konrad-Martin *et al.*, 2001; Mauermann and Kollmeier, 2004). Differences in the level of the reflection component may underlie this variability. One would expect to need a higher level suppressor to suppress a more intense reflection component.

The addition of a suppressor tone does not always lead to a reduction in the depth of the DPOAE fine structure. When the reflection component is larger than the generator component, the fine structure can be enhanced by a suppressor tone. Since the depth of the fine structure is greatest when the two components are equal in level, reduction of the level of a more intense reflection component (such as seen in L6 in Fig. 8) makes the two components more equal in level and thus enhances the fine structure. Additional increases in suppressor level further reduce the level of the reflection component making the difference in level greater, leading to a reduction of the fine structure. If the phase of the DPOAE is evaluated, one can predict if and when the fine structure will increase for each subject as discussed for the 8 s/octave sweeps in experiment 1 (Dhar and Shaffer, 2004; Talmadge *et al.*, 1999). Unfortunately, many investigators do not consider the DPOAE phase. A similar variability of the pattern of change in DPOAE fine structure is seen with noise exposure (Reuter *et al.*, 2007), and when contralateral noise is used to evoke efferent effects (Zhang *et al.*, 2007). It is anticipated that if these investigators had looked at the pattern of phase change with frequency with sufficient frequency resolution to avoid phase wrapping errors, these authors would have been able to predict whether they would see an increase or a decrease in the depth of the fine structure.

Many investigators report the DPOAE level data and do not consider the phase data. A full understanding of the DPOAE depends on looking at both the amplitude and phase. In addition, the phase is much less sensitive to probe position and signal to noise ratio than the DPOAE amplitude. Even when the DPOAE level is in the noise floor, the phase is often stable and varies systematically with frequency, permitting the detection of DPOAE even when they are apparently in the noise floor when phase is not taken into consideration.

IV. DISCUSSION

When the sweep rate is slow enough that the reflection component falls outside the analysis window, using continuously sweeping tones in combination with a LSF analysis gives detailed DPOAE measurements that are almost identical to estimates obtained using discrete DPOAE measurements. The faster sweep provided an efficient procedure for obtaining estimates of the generator component uncontaminated by the reflection component. It is still not clear whether removing the reflection component improves the capacity of DPOAE to predict hearing loss; when the addition of fixed levels of suppressor tones was used on a small number of hearing impaired subjects (Dhar and Shaffer, 2004) or a larger numbers of normal and hearing impaired ears (Johnson *et al.*, 2007), no improvement in diagnostic capacity was seen. However, when Johnson *et al.* (2007) looked at the fine structure in a subset of subjects most likely to be misdiagnosed, they determined that the suppressor levels being used were not always optimal. Furthermore, for this subset of subjects the lowest level suppressor used (which is the only one that did not reduce the magnitude of the generator component), there was an advantage in using the generator alone. The continuously sweeping tone procedure provides an efficient procedure for extracting the generator component alone and does not depend on the use of suppressors. The ability of this procedure to correctly diagnose hearing loss needs to be evaluated.

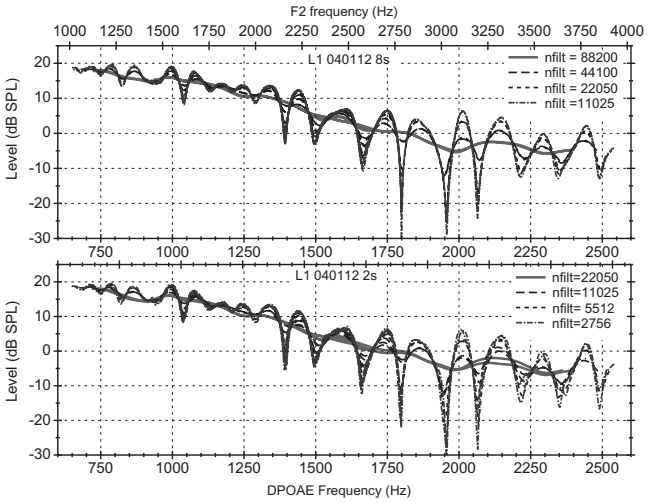


FIG. 10. Estimates of DPOAE level as a function of frequency obtained from 8 s/octave (top) and 2 s/octave sweeps (bottom) when the analysis window (nfilt) is varied. Note that the bandwidth of the analysis filter is determined by dividing the sample rate (41 000 Hz) by the size of the analysis window.

The data presented so far in the paper were all obtained with a fixed analysis window (stepped through the sound file). Since the bandwidth of the analysis depends on the duration of the analysis window, modifying the duration of the analysis window will permit a different number of components to be detected from a fixed sweep rate. Changing the analysis window and thus the bandwidth of the analysis determines the probability that the reflection component will return in time to be included in the analysis window.

The effect of varying the analysis window can be seen in Fig. 10. The top panel shows the DPOAE level estimates from the 8 s/octave sweeps from one subject. The bottom panel shows DPOAE level estimates from the 2 s/octave sweeps from the same subject. The different lines on each plot were obtained by using different analysis windows (nfilt was the number of points used for the analyses) on the same

data. The bandwidth of the analysis filter can be estimated by dividing the sampling rate (44,100) by the size of the analysis window. Note that the pattern of the data is very similar in both panels. The difference is the size of the analysis window (filter bandwidth) needed to obtain a specific pattern. While the smooth generator component pattern is seen in the 2 s/octave sweeps (bottom panel) with the window of 22 050 (2 Hz bandwidth) used in experiments 1 and 2, this pattern is seen with a window of 88 200 (0.5 Hz bandwidth) when the sweep rate is 8 s/octave. The sweep rate is increased by 4 so an equivalent increase in the analysis window is needed. For simplicity, the 16 s/octave data were not included, but a very similar pattern can be seen for this data when the bandwidth is 0.25 Hz.

The depth of the fine structure is increased in both panels when the analysis window was reduced (bandwidth increased). A fine structure equivalent to that seen with the fixed-frequency primaries is seen when the analysis window is 2756 points (16 Hz bandwidth) when analyzing the 2 s/octave sweeps and 11 025 (4 Hz bandwidth) when the 8 s/octave sweeps are being analyzed. The 22 050 analysis window provided the same pattern for the 16 s/octave sweeps in experiment 1. The similarity of the plots in the two panels (despite that they stem from separate sound files obtained several minutes apart) reveals that one does not need to use the slower sweep rates in order to get DPOAE fine structure or more rapid sweeps to obtain estimates of the generator component alone. One can obtain estimates of both DPOAE fine structure or extract the generator component alone from the very efficient 2 s/octave sweeps. As the sweep rate increases, there is, however, some increase in the vulnerability of the DPOAE estimate to noise. An increase in the number of sweeps that are averaged prior to analysis can compensate for this.

The ability to rapidly obtain estimates of DPOAE over a wide range of frequencies, combined with the ability to look at the data in ways that permit the examination of both the DPOAE fine structure and the generator component alone, can be very powerful. A better understanding of the nature of DPOAE fine structure in normal and hearing impaired subjects and across development may improve our understanding of DPOAE generation, changes in cochlea function, and may lead to improved diagnostic tools.

In the experiments reported in this paper, we used overlapping linear sweeps. This is time consuming. A single logarithmic sweep (which leads to an approximation of constant basilar membrane distance/s) is more efficient. We are now obtaining data using this procedure.

V. SUMMARY

Continuously sweeping primaries provides both an efficient procedure for obtaining estimates of DPOAE over a wide frequency range, which permits the extraction of details of the DPOAE fine structure amplitude and phase. Varying either the sweep rate or analysis window permits the extraction of the generator component uncontaminated by later components such as the reflection components.

ACKNOWLEDGMENTS

This work was funded by Grant No. H133E03006 to Gallaudet University from the National Institute on Disability and Rehabilitation Research (NIDRR) of the U.S. Department of Education. Opinions expressed herein are those of the grantee and do not necessarily reflect those of the Department of Education. Two anonymous reviewers provided invaluable comments.

[1]Although group delay (computed using phase derivatives) and signal ("physical") latency are often used interchangeably, it is important to note that the two are only equivalent in a linear, nondispersive system. However, the cochlea is neither linear nor nondispersive, so one would not automatically expect such an association between the two. In particular, for the group delay associated with the generator site typically has nearly a zero latency when obtained using phase derivatives, whereas the physical latency depends on the forward transmission time of at the frequencies of the primaries and the reverse transmission at the DPOAE frequencies. See Tubis *et al.* (2000) for more detailed discussion of this issue.

Dhar, S., and Shaffer, L. A. (**2004**), "Effects of a suppressor tone on distortion product otoacoustic emissions fine structure: Why a universal suppressor level is not a practical solution to obtaining single-generator DP-grams," Ear Hear. **25**, 573–585.

Dhar, S., Talmadge, C. L., Long, G. R., and Tubis, A. (**2002**), "Multiple internal reflections in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

Heitmann, J., Waldmann, B., and Plinkert, P. K. (**1996**), "Limitations in the use of distortion product otoacoustic emissions in objective audiometry as the result of fine structure," Eur. Arch. Otorhinolaryngol. **253**, 167–171.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (**2006**), "Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. **119**, 418–428.

Johnson, T. A., Neely, S. T., Kopun, J. G., Dierking, D. M., Tan, H., Converse, C., Kennedy, E., and Gorga, M. P. (**2007**), "Distortion product otoacoustic emissions: Cochlear-source contributions and clinical test performance," J. Acoust. Soc. Am. **122**, 3539–3553.

Johnson, T. A., Neely, S. T., Kopun, J. G., and Gorga, M. P. (**2006**), "Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. **119**, 3896–3907.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., Cyr, E., and Gorga, M. P. (**2002**), "Sources of DPOAEs revealed by suppression experiments, inverse fast Fourier transforms, and SFOAEs in impaired ears," J. Acoust. Soc. Am. **111**, 1800–1809.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**), "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Long, G. R., and Talmadge, C. L. (**1997**), "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831–2848.

Mauermann, M., and Kollmeier, B. (**2004**), "Distortion product otoacoustic emission (DPOAE) input/output functions and the influence of the second DPOAE Source," J. Acoust. Soc. Am. **116**, 2199–2212.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999**), "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Plinkert, P. K., Heitmann, J., and Waldmann, B. (**1997**), "[Single generator distortion products (DPOAE). Precise measurements of distortion product otoacoustic emissions by three-tone stimulations]," HNO **45**, 909–914.

Reuter, K., Ordonez, R., and Hammershoi, D. (**2007**), "Overexposure effects of a 1-kHz tone on the distortion product otoacoustic emission in humans," J. Acoust. Soc. Am. **122**, 378–386.

Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**), "Sources and mechanisms of DPOAE generation: implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.

Shera, C. A., and Zweig, G. (**1993**), "Noninvasive measurement of the cochlear traveling-wave ratio," J. Acoust. Soc. Am. **93**, 3333–3352.

Siegel, J. H. (**2007**), "Calibrating otoacoustic emission probes," in *Otoacoustic Emissions: Clinical Applications*, 3rd ed., edited by M. Robinette and T. Glattke (Theime, New York), pp. 403–427.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1996**), "Latency and multiple sources of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **99**, 1016–1024.

Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

Tubis, A., Talmadge, C. L., Tong, C., and Dhar, S. (**2000**), "On the relationships between the fixed-f1, fixed-f2, and fixed-ratio phase derivatives of the 2f1-f2 distortion product otoacoustic emission," J. Acoust. Soc. Am. **108**, 1772–1785.

Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**), "Generation of DPOAEs in the guinea pig," Hear. Res. **178**, 106–117.

Zhang, F., Boettcher, F. A., and Sun, X. M. (**2007**), "Contralateral suppression of distortion product otoacoustic emissions: effect of the primary frequency in Dpgrams," Int. J. Audiol. **46**, 187–195.

Spatial release from masking with noise-vocoded speech

Richard L. Freyman,[a] Uma Balakrishnan, and Karen S. Helfer
Department of Communication Disorders, University of Massachusetts, 358 North Pleasant Street, Amherst, Massachusetts 01003

(Received 19 June 2007; revised 4 June 2008; accepted 5 June 2008)

This study investigated how confusability between target and masking utterances affects the masking release achieved through spatial separation. Important distinguishing characteristics between competing voices were removed by processing speech with six-channel envelope vocoding, which simulates some aspects of listening with a cochlear implant. In the first experiment, vocoded target nonsense sentences were presented against two-talker vocoded maskers in conditions that provide different spatial impressions but not reliable cues that lead to traditional release from masking. Surprisingly, no benefit of spatial separation was found. The absence of spatial release was hypothesized to be the result of the highly positive target-to-masker ratios necessary to understand vocoded speech, which may have been sufficient to reduce confusability. In experiment 2, words excised from the vocoded nonsense sentences were presented against the same vocoded two-talker masker in a four-alternative forced-choice detection paradigm where threshold performance was achieved at negative target-to-masker ratios. Here, the spatial release from masking was more than 20 dB. The results suggest the importance of signal-to-noise ratio in the observation of "informational" masking and indicate that careful attention should be paid to this type of masking as implant processing improves and listeners begin to achieve success in poorer listening environments. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2951964]

PACS number(s): 43.66.Dc, 43.66.Qp, 43.66.Pn, 43.66.Ts [AJO] Pages: 1627–1637

I. INTRODUCTION

When a target sound is presented together with a masking sound, similarities between the two sounds can create confusion that interferes with the detection and recognition of the target. Evidence for this type of interference caused by target/masker confusability is found where increased thresholds or poor discrimination cannot be explained by traditional conceptualizations of masking. The term most commonly applied in these situations is "informational masking," which in recent years has frequently been considered with regard to the masking of speech by competing speech utterances (e.g., Freyman *et al.*, 1999; Brungart *et al.*, 2001; Arbogast *et al.*, 2002; Hawley *et al.*, 2004; Kidd *et al.*, 2005).

Substantial dissimilarity between target and masker is thought to minimize or eliminate informational masking (Durlach *et al.*, 2003). For example, little evidence of informational masking is found with target and masker speech spoken by talkers of the opposite sex (Brungart *et al.*, 2001; Brungart and Simpson, 2002), presumably because large differences in fundamental frequency between male and female voices minimize confusion. Further, different spatial impressions caused by target and masker are likely to reduce confusion between them and severely reduce or eliminate informational masking (e.g., Freyman *et al.*, 2001; Gallun *et al.*, 2005; Kidd *et al.*, 2005).

The two experiments reported in this paper used noise-excited vocoded speech to investigate informational masking under conditions in which target-masker similarity is expected to be high. This type of speech processing, in which envelopes in different frequency channels are extracted and used to modulate a noise carrier, has been used to model key aspects of processing by cochlear implants (e.g., Shannon *et al.*, 1995; Dorman *et al.*, 1998; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Poissant *et al.*, 2006). This processing can severely reduce or eliminate pitch and other voice difference cues between different talkers, increasing target/masker confusability. Using noise-vocoded speech without target/masker spatial differences, both Qin and Oxenham (2003) and Stickney *et al.* (2004) demonstrated more masking from a single competing talker than predicted from the pattern of results obtained with unprocessed speech. Both reports suggest that vocoded speech may be especially susceptible to informational masking. Stickney *et al.* (2007) offered evidence of improved target recognition under some conditions when temporal fine-structure cues sufficient to provide fundamental frequency information were added to noise-vocoded speech.

Paradoxically, increased susceptibility to informational masking with vocoded speech could be partially mitigated by the difficulty listeners have in understanding speech subjected to this kind of processing. In order to achieve reasonable levels of performance, it is sometimes necessary to present the stimuli at high signal-to-noise (SN) ratios of 10 dB or more (e.g., Qin and Oxenham, 2003; Poissant *et al.*, 2006). At these SN ratios, the target is much louder than the masker. Although partial masking may obscure lower level portions of the waveform and may lead to reduced intelligibility, at these positive SN ratios there should be no overall confusion between target and masker. In fact, Arbogast *et al.* (2005) hypothesized that informational masking may decline

[a]Author to whom correspondence should be addressed. Tel.: 413-545-0298. Electronic mail: rlf@comdis.umass.edu

dramatically above 0 dB SN ratio. However, even though a vocoded target at high SN ratios may stand out because it is louder than interfering speech, the lower level portions of the speech-envelope-modulated noise belonging to the target may be confused with the modulations in the masker (Qin and Oxenham, 2003).

The goal of the present study was to understand the influence of SN ratio on informational masking with vocoded speech. The first experiment examined the recognition of open set nonsense sentences in a background of two-talker masking, where positive SN ratios were required for above-floor performance. The second experiment used the same masking stimuli and a subset of the target stimuli, but there the task was only to *detect* the presence of target stimuli in an adaptive forced-choice task. Negative SN ratios were sufficient for threshold performance in this task. Thus, across the two experiments, identically processed target and masking stimuli were employed at very different SN ratios, giving us insight into how informational masking was influenced by SN ratio. In both experiments, informational masking was quantified by measuring spatial release from masking under two-source speech masking conditions that produce no release from continuous noise masking.

II. EXPERIMENT 1: OPEN-SET NONSENSE SENTENCE RECOGNITION

A. Methods

1. Stimuli

The target stimuli were a set of 320 "nonsense" sentences that were syntactically but not semantically correct, e.g., "A *shop* can *frame* a dog." Each sentence included three key words, as underlined in the example. These sentences, recorded by a female talker, have been used in several earlier studies (e.g., Helfer, 1997; Freyman *et al.*, 1999, 2001, 2007; Li *et al.*, 2004). A full description of the recording methodology can be found in Helfer (1997) or Freyman *et al.* (1999). The maskers were nonsense sentences recorded by ten different female talkers (different sets for different talkers). Details of the recording methodology for these maskers can be found in Freyman *et al.* (2007). Pauses were removed between sentences, creating an approximately 35 s long stream for each talker. The streams were equated in average power (rms) and then combined to form five two-talker maskers, with the selection of pairings roughly according to average fundamental frequency.

The target and masking stimuli were processed through six-channel vocoding with a noise carrier using the same algorithm as that in Qin and Oxenham (2003). The frequency range from 80 to 6000 Hz was divided into six channels of equal bandwidth according to the equivalent rectangular bandwidth (ERB) scale (Glasberg and Moore, 1990), using digital sixth-order butterworth bandpass filters. Envelopes were extracted from the filter outputs by digitally low-pass filtering rectified signals with a cutoff frequency of the larger of 300 Hz or half the bandwidth, using a second-order butterworth filter. White noise filtered to have the same bandwidth as the filtered signals was multiplied by the appropriate envelope channel in the time domain to create noises that matched the temporal envelopes in each channel. The six modulated noises were summed to create a broadband six-channel speech-envelope-modulated noise for each of the 320 target sentences and the five two-talker masker speech streams. This type of modulated noise has been shown in several studies to be quite intelligible for sentence stimuli with just four channels (Shannon *et al.*, 1995).

2. Environment and apparatus

The experiment was conducted in a large double-walled sound-treated room (IAC No. 1604) measuring 2.76 $\times$2.55 m. Reverberation times measured in this room ranged from 0.12 s at 6.3 and 8.0 kHz to 0.24 s at 125 Hz (Nerbonne *et al.*, 1983). A previous study conducted in this room using the target front and masker right front (F-RF) condition (Helfer and Freyman, 2005) showed the same kinds of spatial release that have been found in an anechoic chamber (e.g., Freyman *et al.*, 2001). The listener sat on a chair placed with its back against one wall of the room. Two loudspeakers (Realistic Minimus 7), at a distance of 1.3 m from the center of the head when seated in a chair and a height of 1.2 m (the approximate height of the ears of a typical listener), delivered the target and masking stimuli. One loudspeaker was placed at 0 deg azimuth, directly in front of the listener; the second loudspeaker was at 60 deg to the right. The target and masking stimuli were mixed digitally at the appropriate SN ratio on a computer before presentation from two channels of the computer's sound board, attenuated (TDT PA4), amplified (TDT HBUF5), power amplified (TOA P75D), and delivered to the loudspeakers.

In the front-front (F-F) condition, target and masker were presented from the front loudspeaker. In the F-RF condition, the target was presented from the front loudspeaker and the masker from both loudspeakers, with the right leading the front by 4 ms. Due to the precedence effect, the F-RF masking configuration creates the perception that the masker is to the right, well separated from the front target. The 4 ms delay version of this F-RF configuration has been shown to create little or no release from masking for speech targets using continuous or fluctuating noise maskers, indicating no energetic masking release (Freyman *et al.*, 1999; Brungart *et al.*, 2005; Rakerd *et al.*, 2006). Further, Helfer and Freyman (2005) demonstrated no release from continuous noise masking for this configuration in the same sound-treated room used for the current studies. It is assumed, therefore, that when masking release occurs for the F-RF configuration in a competing speech task, the release from masking is due to nonenergetic effects, where perceptual spatial differences make it easier to extract the target from the complex mixture of voices.

3. Subjects

Listeners were ten adult students with hearing thresholds $\leq$20 dB HL in the frequency range of 500–4000 Hz (ANSI S3.6, 1996). The ages ranged from 19 to 21 years. None had extensive experience listening to vocoded speech.

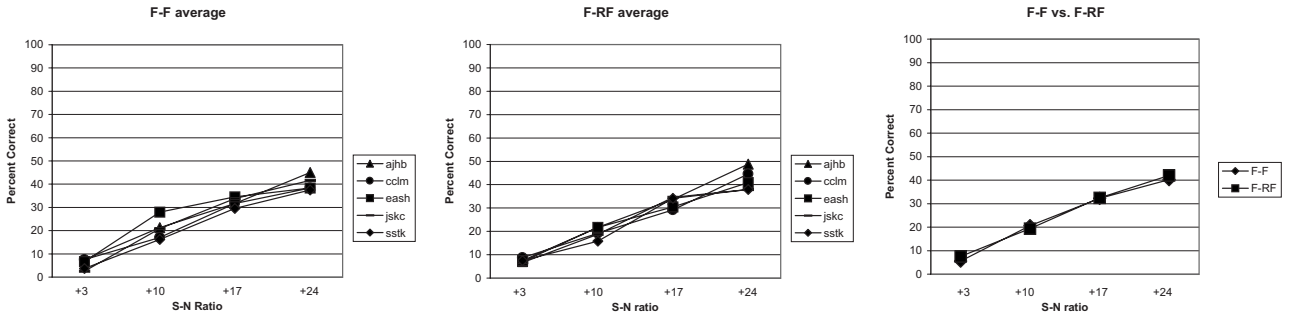


FIG. 1. Percent correct recognition of key words as a function of SN ratio for five two-talker maskers. The left panel shows data for the F-F configuration and the middle panel for the F-RF configuration. The right panel shows F-F and F-RF data combined across all five maskers.

4. Procedures

Subjects were seated in a chair and instructed to face the front loudspeaker but were not physically restrained. At the beginning of each trial, the word "ready" appeared on a computer screen in front of the subject. The presentation of the masker was then initiated, followed by the presentation of the target 0.6–1.2 s later. The masker was terminated simultaneously with the end of the target sentence. The listener repeated the target sentence out loud, and the experimenter, monitoring in a control room, scored the three key words as correct or incorrect.

On each trial a section of two-talker masker waveform was selected randomly from the 35 s stream. The masker onset could occur anywhere in the stream, e.g., at the beginning of a sentence for one of the two talkers and in the middle for the other talker. As noted above, the target sentence began 0.6–1.2 s after the beginning of the masker. Subjects were told that they could use this as a cue for directing attention to the target speech in the presence of the masker.

A total of 40 conditions were presented to each of the listeners: five two-talker maskers, two spatial configurations (F-F and F-RF), and four SN ratios chosen during pilot listening (+3, +10, +17, and +24 dB), presented in a completely crossed design. SN ratio was defined as target rms amplitude relative to masker rms amplitude. SN ratios were manipulated by changing the level of the masker for a fixed-level target, which was always presented at 44 dBA (calibrated at the position of the subject's head using a speech spectrum noise with the same rms as the target sentences). The 320 target sentences were selected at random, without replacement, and with a different random order for each listener. The sentences were presented in 40 blocks of eight sentences each. The SN ratio, masking talkers, and spatial configuration were all fixed within a block. Across sets of four blocks (32 trials), only the SN ratio changed (randomly), while the spatial configuration and masker were fixed. A second set of four blocks followed where the spatial configuration switched from F-F to F-RF or vice versa. Half the subjects received F-F first and the other half F-RF first. After eight blocks were presented (64 trials), the masker changed and the process was repeated until all five maskers had been presented for 64 trials. The order of presentation of the five maskers within a set of 320 trials was random and different for each listener. Subjects completed the experiment in one listening session.

Prior to actual data collection, subjects listened to five practice examples of vocoded nonsense sentences. Each sentence was presented several times in quiet. If the listener did not respond correctly after two or three repetitions, the unprocessed version of the target was presented, and then the vocoded sentence was repeated again until the subject was able to recognize the vocoded sentence.

B. Results and discussion

Results of experiment 1, displayed in Fig. 1, show that even at high SN ratios only moderate levels of performance were obtained. Performance with the same six-channel processing was found to be much better with meaningful sentences (Poissant *et al.*, 2006), so the poorer results probably reflect the use of difficult nonsense sentences where there is no semantic context. The data in the left and middle panels show limited variability in masking efficiency across the five different maskers; in contrast a substantial variation was seen across the same five maskers in their unprocessed version, particularly for the F-F condition (Freyman *et al.*, 2007). For example, at the second highest SN ratio used in that study, performance ranged from 18% correct with one two-talker masker to 71% correct with a different two-talker masker. To be sure, the upper range of performance in the current study may be constrained by the intrinsic difficulty of the stimuli even in the absence of masking. However, the striking homogeneity of performance across masking talkers suggests that some aspect of voice pitch or quality lost in the vocoding process accounted for the variability observed with the unprocessed stimuli in Freyman *et al.* (2007). It suggests little variation in masking produced by other features of the maskers that were preserved in the vocoding process, such as gross temporal patterns. To the extent that vocoding simulates listening with a cochlear implant (e.g., Shannon *et al.*, 1995; Dorman *et al.*, 1998; Qin and Oxenham, 2003; Stickney *et al.*, 2004; Poissant *et al.*, 2006), one might expect to find reduced variability in performance across masking talkers in competing speech situations among cochlear implant users. As an example, Stickney *et al.* (2004) did not find statistically significant differences in implanted listeners' performance in the presence of three different masking talkers,

although the same conditions delivered to normal-hearing listeners showed significant variability across masking talkers.

Perhaps the most important result, shown in the right panel, is that there was no benefit of presenting the target and maskers in the spatial (F-RF) configuration. The same manipulation improved performance by as much as 30 percentage points with the unprocessed stimuli in Freyman *et al.* (2007), representing a 6 dB shift in SN ratio for equivalent (50% correct) performance. The lack of improvement with the current stimuli seems exactly the opposite of what might have been expected. That is, as noted in the Introduction, it would not have been surprising to observe a great deal of informational masking in the F-F configuration because of the absence of quality and pitch differences between the target and maskers in the vocoded tokens. If this type of interference was released by the F-RF presentation of the masker, the improvement could have been quite large. At least three explanations can be considered as to why no advantage was found.

First, it is possible that informational masking was present in the F-F configuration but was not released in the F-RF configuration because the latter condition did not produce different spatial perceptions for the target and masker. Informal observations by the experimenters and four naïve listeners who marked perceived spatial positions on a photocopy of a protractor suggested that this explanation was not correct. The vocoding process did not interfere with localization; i.e., the masker was heard well to the right in the F-RF configuration, and the target appeared from the front loudspeaker.

A second possibility is that a vocoded masker does not, in general, produce informational masking with a vocoded target, despite expectations to the contrary as explained above. When both the masker and target are impoverished in fine structure, leading to decreased intelligibility, the effectiveness of the interfering talkers in producing informational masking could have been theoretically reduced. On the other hand, the results of a number of studies (e.g., Arbogast *et al.*, 2002, 2005; Gallun *et al.*, 2005; Shinn-Cunningham *et al.*, 2005) using vocoded spectrally interleaved targets and maskers indicate that informational masking can be quite strong when the stimuli are spectrally degraded.

The third and, we believe, most likely reason that evidence of informational masking was not seen was the high SN ratios needed in this experiment. The combined use of nonsense sentence materials and envelope vocoding required SN ratios to be very high—so high that the target was clearly distinguishable from the maskers. That is, the difference in loudness between the target and masker probably reduced the confusability between them and limited the interference to purely energetic masking. Indeed, such was the impression of the investigators listening to the materials. If we assume that loudness differences reduce confusability, the greatest chance of seeing a benefit of the F-RF configuration would be where the intensity of the target and masker was most comparable (+3 dB SN ratio). However, even that SN ratio may be high enough to overcome informational masking, as the target is 6 dB above the level of either of the individual maskers in the two-talker complex. Also, the task at that SN ratio may have been too difficult for a benefit to be seen due to a floor effect.

The finding of no indications of informational masking in experiment 1 is an interesting result because it provides a counterexample to the results of Qin and Oxenham (2003) and Stickney *et al.* (2004), both of which seem to indicate increased informational masking with vocoded targets and maskers. The difference could be related to our use of a combination of two masking talkers rather than the one masking talker used in the earlier studies. Also, in Qin and Oxenham (2003) and Stickney *et al.* (2004), the indications of informational masking come from increased masking with speech maskers relative to continuous noise maskers. The method of assessing informational masking in our study is to measure spatial release from masking for conditions where release from energetic masking is not expected. In the second experiment we asked whether spatial release from masking could be observed with the processed targets and maskers using a different task that required lower SN ratios.

III. EXPERIMENT 2: DETECTION THRESHOLDS

When both target and masker are processed with the same vocoding and presented from the same (front) loudspeaker, it is likely that many of the cues necessary for extracting the target from the interference are absent. In experiment 1 the SN ratio was high enough so that the target may have stood out against the less intense masker, minimizing target/masker confusion and eliminating the benefits of providing spatial differences between the target and masker. In order to evaluate the hypothesis that the high SN ratios used in the recognition study were responsible for the absence of spatial release from masking, we sought to create stimulus conditions that were similar but where the task could be performed at lower SN ratios. In this experiment we used words excised from the target sentences, one of the five maskers from experiment 1, and the identical vocoding process. The primary difference was that the task for the subject was only to detect the presence of the target words in a four-interval four-alternative forced-choice (4AFC) adaptive paradigm. Assuming that the identical processing for the target and masker could create confusion that would affect even the detection of the presence of the target, alleviation of this confusion through the introduction of spatial differences could potentially lead to the kind of spatial release from masking expected, but not seen, in experiment 1.

A. Methods

1. Stimuli

Target stimuli were 20 consonant-vowel-consonant (CVC) words excised from the nonsense sentences used in experiment 1. Details described below about word selection and excision are identical to those in Balakrishnan and Freyman (2008). The 20 target words were chosen for clarity of production and ease of excision from the nonsense sentence waveforms. Typically, the target word was the second of the three keywords of each utterance. In those instances where the second word in the sentence was not easy to extract, first

or third keywords were taken. The 20 target words were either processed with the same six-channel vocoding used in experiment 1, or they were left unprocessed. They were scaled to equate their average power (rms) and then postpadded with zeroes to match the duration of longest word in the list (500 ms). The 20 words were concatenated to create a single file from which the experimental software randomly selected a single word and played it on each trial.

The majority of conditions employed a two-talker speech masker (sstk), one of the five maskers used in experiment 1. The unprocessed version was used with the unprocessed target words, and the six-channel vocoded version was used as a masker for the vocoded targets. An additional masker, speech spectrum noise (Byrne *et al.*, 1994), was used with three listeners to verify earlier findings in recognition experiments that spatial separation produced no release from purely energetic masking.

2. Apparatus

The detection experiments were conducted in an anechoic chamber measuring $4.9 \times 4.1 \times 3.12$ m. The walls, floor, and ceiling are lined with 0.72 m foam wedges. Subjects were seated in the center of the room in front of a foam-covered semicircular arc on which two loudspeakers were positioned. The front loudspeaker was at 0 deg horizontal azimuth; the right loudspeaker was at 60 deg to the right. Both were 1.9 m from the approximate center of the subjects' head and were at ear height for the typical adult.

The target words were delivered via TDT System I instrumentation. The output of the 16 bit digital to analog converter (TDT DA1) running at 20 kHz was low-pass filtered at 8.5 kHz (TDT), attenuated (TDT PA3), and mixed with the masker before being delivered to a Crown D40 amplifier and a Realistic Minimus 7 loudspeaker. The masker was delivered from a second computer (Dell Dimension XPD 333) via audio software (COOL EDIT PRO). The 35 s long interference segment was played continuously in loop mode over the duration of an adaptive track. The masker was attenuated (TDT PA4) before being mixed with the target. Calibration of the target and maskers was completed by measuring sound levels at the position of the subject's head with the subject absent. The target was calibrated to a sawtooth noise equated for average power to the target words. The maskers were calibrated for each channel using the speech-shaped noise masker.

3. Subjects

Listeners were five college students with hearing thresholds $\leqslant$20 dB HL in the frequency range of 500–4000 Hz (ANSI S3.6, 1996). The ages ranged from 19 to 43 years with a median of 21 years. None of the subjects participated in experiment 1, and none were well practiced in listening to vocoded signals.

4. Conditions and procedures

The target words always originated from the front loudspeaker. The same F-F and F-RF masking configurations from experiment 1 were used in this experiment also. In addition, a target front and masker front right (F-FR) condition was used with the two-talker maskers, where the masker from the front led the masker from the right by 4 ms, the opposite of the F-RF configuration. Thus, for a given run the target-masker configuration could be F-F, F-RF, or F-FR.

Subjects responded to each 4AFC trial using a button box with LEDs that marked the four intervals, one of which, selected randomly with equal probability, contained the target. Listeners were instructed to indicate the interval in which they thought they heard the target. Other than informing them that the target was always from the front, no special instructions were given to participants on how they might solve the task. Feedback was provided via an LED display that illuminated the target interval. For all conditions, the masker level was fixed at 53 dBC in each masker channel while the target level was adapted. A two-down one-up stepping rule was employed to estimate 70.7% criterion performance (Levitt, 1971). An individual adaptive track consisted of ten reversals, with the threshold computed as the arithmetic mean of the last six reversals. The initial step size for the adaptive track was 16 dB, which was halved after each reversal until a final step size of 2 dB was reached. For each condition, three consecutive adaptive tracks were run and final threshold determined as the arithmetic mean of the three individual thresholds.

At the beginning of the first listening session, subjects were verbally instructed, familiarized with the list of 20 target words (via print and audio), and given practice runs in quiet to familiarize them with the task. Before the main experiment was begun, one adaptive threshold estimate was obtained in quiet for both the unprocessed and vocoded targets. These were no higher than 6 dBC for both processing conditions. The sequence of runs for the main experiment was as follows.

First, data were collected for the F-F and F-RF configurations for both the vocoded and unprocessed stimulus conditions. The order of the collection of these four thresholds was randomly determined for each listener.

Second, all five subjects participated in a brief investigation of potential learning effects. A previous study with unprocessed target and masking stimuli (Balakrishnan and Freyman, 2008) had shown that some of the listeners' performance improved (albeit erratically) over successive exposure to the F-F condition. In the current experiment, learning effects were studied by obtaining ten additional adaptive runs in the F-F configuration for both the vocoded and unprocessed stimulus conditions. The order of the processing conditions was counterbalanced as best as possible with five listeners. The ten runs for each processing condition were run consecutively in a block with a break after the first five runs.

Third, thresholds were obtained for the F-FR configuration, with a random ordering of vocoded and unprocessed conditions across subjects. Unlike the F-RF condition, the F-FR configuration does not provide a large angular separation of the target and masker. This is because the lead sound in the two-source masker emanates from the same location (front) as the target. Several speech recognition studies using unprocessed target and maskers (Freyman *et al.*, 1999; Brun-

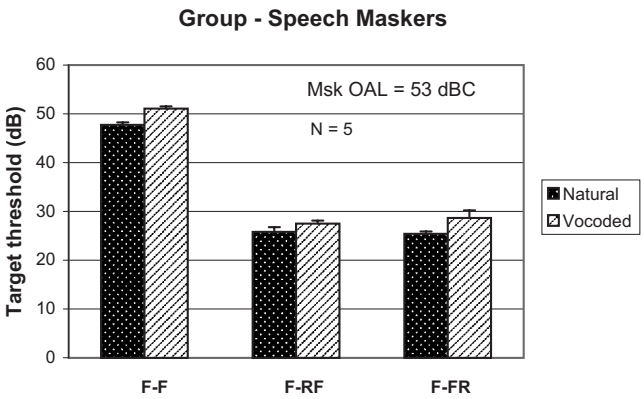


FIG. 2. Across-subject mean masked detection thresholds (in dBC) for CVC words excised from the nonsense sentences used in experiment 1. Thresholds for the F-F, F-RF, and F-FR configurations are shown for both unprocessed and vocoded speech. Error bars represent one standard error of the mean.

gart *et al.*, 2005; Rakerd *et al.*, 2006) and one detection study (Balakrishnan and Freyman, 2008) had shown that wide angular separations were not essential for release from informational masking. In the present study, we examined whether informational masking release would occur in the F-FR condition for the degraded and presumably more confusable vocoded stimuli.

Finally, three of the listeners also completed three adaptive tracks with a speech-shaped noise masker (Byrne *et al.*, 1994) for the F-F and F-RF loudspeaker configurations. Data were collected for only a subset of listeners for this masker because prior data collected for speech targets in noise showed no benefits of perceived spatial separation (Freyman *et al.*, 1999).

B. Results

Figure 2 displays mean detection thresholds and ±1 standard error for the nonspatial (F-F) and spatial (F-FR and F-RF) target/masker configurations. For the F-F configuration, detection thresholds averaged approximately 48 and 51 dBC for the unprocessed and vocoded speech, respectively. These correspond to SN ratios of −5 and −2 dB. Each spatial masking configuration produced reductions in detection thresholds of approximately equivalent sizes for both unprocessed and vocoded speech. The reductions, at 20 dB or greater, were considerably larger than the 5–8 dB spatial release from masking observed with recognition of the full unprocessed sentence stimuli and the same F-RF maskers in previous studies (Freyman *et al.*, 2001, 2007). The amount of release from masking is comparable to that observed for a different group of untrained subjects who listened to the unprocessed targets and maskers in a separate study (Balakrishnan and Freyman, 2008). The large spatial release could be due to the fact that in the F-RF configuration, the listener only had to detect the presence of any sound from the front loudspeaker, which required very low SN ratios in the fluctuating masker. On the other hand, in the F-F condition, merely detecting the presence of sound from the front loudspeaker was obviously not sufficient, and much higher levels were necessary for the target to stand out from the background in one of the intervals.

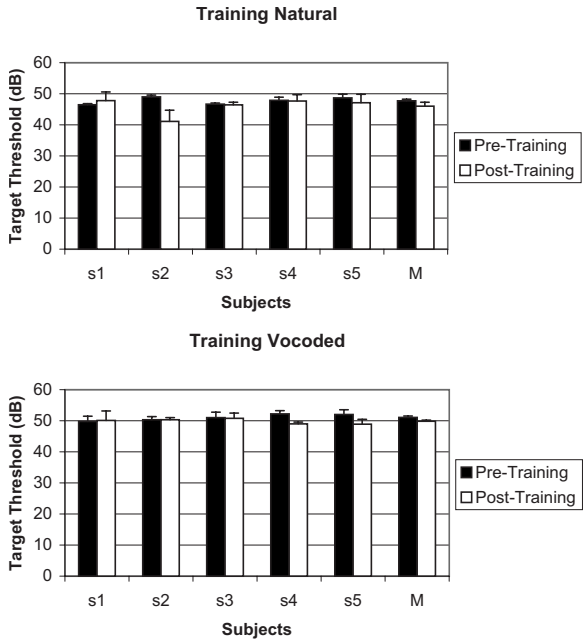


FIG. 3. Individual-subject and mean (*M*) masked detection thresholds (in dBC) for the words in the F-F configuration before and after extended listening experience. The solid bars are the average of the initial three runs, with the across-subject mean reproduced from Fig. 2. The open bars are the average of the last three of ten consecutive additional runs on the same condition. The top panel displays the data for the unprocessed conditions, and the bottom panel those for the vocoded conditions. Error bars for the individual subject data represent the standard deviation across three runs, while the error bars for the mean thresholds are one standard error.

Individual detection thresholds for the average of the last three of training runs completed for the F-F condition are shown in Fig. 3. They are compared with the average of the initial three runs for the same condition (the across-subject mean is a replot of data from Fig. 2). From the unprocessed data (top panel) it is clear that four of the subjects show pre- and post-training thresholds that are within error of measurement; a fifth subject (S2) showed an 8 dB improvement post-raining but with high variability across runs. The mean difference across subjects between pre-and post-training thresholds was not statistically significant ($t=1.064$, $p=0.347$). A few of the listeners in Balakrishnan and Freyman (2008) also had shown improvements for unprocessed targets and maskers in the post versus pretraining runs, but once again with high variability across runs. As shown in the bottom panel of Fig. 3, for all five listeners performance with the vocoded stimuli was relatively unaffected by the amount of experience acquired in the present study, and the mean difference between pre- and post-training thresholds was not statistically significant ($t=1.576$, $p=0.190$).

The results of the control condition when the masker was steady-state noise are shown in Fig. 4. In the recognition task (e.g., Freyman *et al.*, 2001, Helfer and Freyman, 2005; Brungart *et al.*, 2005; Rakerd *et al.*, 2006), the F-RF configuration with noise masking and a 4 ms delay provides no reliable improvement relative to the F-F condition. The data

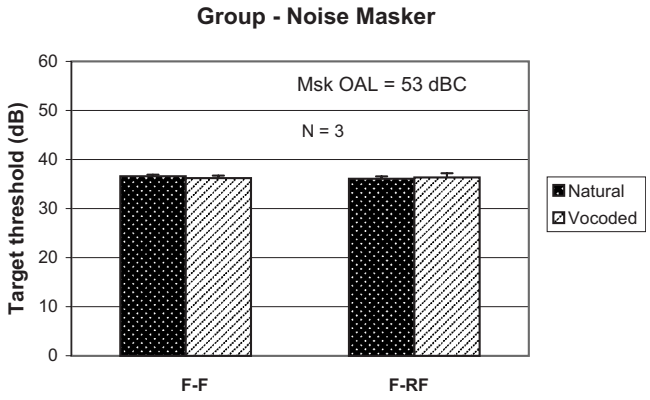


FIG. 4. Across-subject mean detection thresholds for target words (in dBC) in the presence of the noise maskers. Thresholds are shown for both vocoded and unprocessed conditions in both the F-F and F-RF configurations. Error bars represent one standard error of the mean.

in Fig. 4 confirm this lack of improvement in the current detection task for both unprocessed and vocoded speech.

IV. DISCUSSION

A. Summary of results

(1) *A priori* expectations to the contrary, no release from masking was found for recognition of six-channel noise-vocoded nonsense sentences using a spatial configuration that has been shown to produce a large release from informational masking with unprocessed stimuli.
(2) Variability in masking effectiveness across five different pairs of masking talkers was extremely small with the vocoded stimuli, in contrast to the observation of considerably larger variability with unprocessed versions of the same targets and maskers (Freyman *et al.*, 2007).
(3) A 4AFC adaptive detection task using words excised from the sentences revealed large amounts of spatial release from presumably informational masking, although not larger than what was observed for unprocessed stimuli. The spatial release was nearly equally effective whether or not there was a large difference in the target/masker perceived angle (F-RF versus F-FR).
(4) Short-term experience in the nonspatial condition through extended repetitions of adaptive runs did not reveal substantial evidence that listeners could learn to overcome masking in a condition where the informational component was presumably large.

B. Interpretation of the recognition data

In experiment 1, no release from masking was found when the vocoded target speech was presented from the front and the vocoded masking speech was presented from both the front and right loudspeakers. This lack of improvement contrasts strongly with the sizeable release from masking found for the same target and masker without vocoding (Freyman *et al.*, 2007). In one sense, the absence of masking release for the vocoded stimuli in the F-RF condition is counterintuitive because in this configuration masking release has been assumed to be related to a reduction in target/masker confusion caused by the spatial differences between the target and masker (Freyman *et al.*, 1999). With noise-excited vocoded speech, the confusion between target and masker could be expected to increase sharply, creating more informational masking and therefore increasing the release of masking that could be realized with spatial differences.

The fact that absolutely no spatial release from masking was found is due, we believe, to the fact that high SN ratios were required for reasonable performance with the vocoded nonsense sentences. At these SN ratios (ranging from +3 to +24 dB), the target speech was louder than the masking speech and should have stood out from the background. It is likely that there was no overall confusability between the target and masking utterances. Still, although the target was clearly audible there were nevertheless steady improvements with increasing SN ratio. This could occur because low-level portions of the target envelope-modulated noise could have been masked by or confused with the higher level portions of the masker envelope-modulated noise, to a decreasing degree as the SN ratio was increased. If confusion, as opposed to only energetic masking, was involved, then a release from masking would have been expected in the spatial conditions. We interpret the absence of spatial release from masking as an indication that there was no informational masking with the vocoded sentences in this experiment. This explanation is in agreement with the data and interpretations of Arbogast *et al.* (2005), who proposed that 0 dB may be near the upper limit of SN ratios where informational masking is likely to be observed.

It may be useful to consider how well the idea of an SN ratio ceiling fits with data from nonspeech informational masking tasks. A most compelling example is found in the data of Oxenham *et al.* (2003). That study looked at the effect of musical training on informational masking. Tones were detected in the presence of two types of multitone complexes that were either likely or not likely to produce informational masking. Nonmusicians were significantly and substantially more susceptible to informational masking with these signals. Most striking from the perspective of the current study is that for listeners showing significant informational masking (nonmusicians), over half of the informationally masked thresholds and, in particular, thresholds for the five listeners with the most informational masking clustered within a few decibels of 0 dB SN ratio; yet virtually no data point was reliably above 0 dB (Oxenham *et al.*, 2003; Fig. 2). The notion of a ceiling in the vicinity of 0 dB appears to be well supported by their data.

The speech recognition results of Qin and Oxenham (2003) may provide a counterexample to the idea that informational masking cannot be observed at positive SN ratios. Target utterances were lists of HINT sentences (Nilsson *et al.*, 1994). Maskers were male and female talkers, modulated noise, and steady speech spectrum noise. For unprocessed sentences, speech reception thresholds (SRTs) were lowest (best) with the single male and female talker maskers. However, when the targets and maskers were vocoded with the same type of processing used in the current study, the single-talker maskers actually produced the *highest* (poorest) thresholds. Most notably, this tendency continued and even increased slightly when the number of envelope channels

was reduced from 24 to 4 and the SRTs were between +10 and +20 dB SN ratio. To the extent that the increased masking efficiency of the speech maskers relative to the noise maskers demonstrates informational masking, these data must be taken as evidence of informational masking at positive SN ratios. Similar results and interpretations were offered by Stickney *et al.* (2004), and the results were also extended to actual cochlear implant listeners.

Although the results of Qin and Oxenham (2003) and Stickney *et al.* (2004) suggest that informational masking may exist in actual and simulated implant listening, even at positive SN ratios, there are some differences from the current study that could be important. First, both studies used single-talker maskers, while the current experiments used two-talker maskers. Poissant *et al.* (2006), also using six-channel vocoded speech, did not find that a two-talker masker was substantially more effective than a speech spectrum noise masker. There could be something fundamentally different about the confusions that occur with single-talker maskers, especially when the target and maskers are essentially modulated noise. Second, the measure of informational masking is different. Both of the studies referred to above discussed their data in terms of the difference in the masking efficiency of speech maskers versus speech spectrum noise, whereas in the current study we measured the spatial release from masking caused by spatial masker presentations that have produced no evidence of release from energetic masking.

One issue that should be considered when interpreting the speech masker versus noise masker comparison is that the net effect of the valleys and peaks of a fluctuating masker may not remain constant with changes in SN ratio. The fluctuations in a speech masker cause spectrotemporal peaks and valleys relative to an unmodulated noise masker of equal rms amplitude. The traditional data indicate that the effect of the additional masker energy in the peaks is more than offset by the reduced energy in the valleys, which allows relatively unmasked glimpses of the target speech (e.g., Festen and Plomp, 1990). However, to our knowledge, this demonstration has been made mostly, and perhaps exclusively, at low SN ratios [see examples in Festen and Plomp (1990) and George *et al.*, (2006)]. At low SN ratios, spectrotemporal peaks may be relatively inefficient, adding energy but not masking if portions of the target are already inaudible. Conversely, at strongly positive SN ratios where much of the speech is already audible, the benefits of listening in masker valleys may be more than offset by the additional masking resulting from the masker peaks.

At least for unprocessed speech, the differences in slope between the psychometric functions obtained with speech versus noise maskers lend support to the idea that improvements with speech maskers are most obvious at low SN ratios and may disappear at higher SN ratios. For example, although the unprocessed data of Qin and Oxenham (2003) show a better SRT (50% correct) for speech maskers than noise maskers, the mean sigmoidal parameters used to fit the psychometric functions suggest that the efficiency of their male speech masker is reduced relative to a speech spectrum noise masker only below −4 dB SN ratio, at which both functions reach 80% correct. Above that, the fits predict slightly *better* performance in the noise masker. The data for the female masker, which should produce little informational masking of the male HINT sentence talker, show a similar trend of converging with the noise masker data (at −2 dB SN ratio). Results from Stickney *et al.* (2004) (Fig. 2) show no difference between noise and speech maskers at 0 dB SN ratio at approximately 80% correct, but a much slower decline in performance in the presence of the speech masker as SN ratio is reduced. Thus, the relative efficiency of speech spectrum noise maskers and single-talker speech maskers is highly dependent on the SN ratio, which should be taken into consideration when interpreting the difference between them in terms of informational masking.

C. Interpretation of the detection data

Our own measure of informational masking, spatial release from masking with a two-source relative to a single-source masker, must also be considered carefully, as there potentially could be other factors besides positive SN ratios to explain why no masking release was measured. We felt it was important to support our interpretation by demonstrating masking release for the processed stimuli at lower SN ratios. In experiment 2, words excised from the sentence stimuli were used in a detection study. The SN ratios required for detection in the spatial masking conditions were all below −20 dB. If informational masking contributed to thresholds in the nonspatial (F-F) condition there would be ample room to observe it before any truncation could occur at 0 dB SN ratio. Indeed, the nonspatial thresholds were much higher than the spatial thresholds for both vocoded and unprocessed speech. They averaged −2 and −5 dB SN ratio respectively, and suggest an exceptionally large amount of informational masking (20–25 dB).

The fact that the spatial/nonspatial difference is larger than what has been observed previously for the recognition of similar unprocessed stimuli (Freyman *et al.*, 2007) may be partially explained by the number of decibels below 0 dB SN ratio at which the spatial thresholds are obtained. *Recognition* of the unprocessed nonsense sentences in the presence of the same spatial two-talker masker became nearly impossible as the SN ratio was reduced to −12 dB (Freyman *et al.*, 2001). Assuming a ceiling in the vicinity of 0 dB SN ratio in the nonspatial condition, it would not be possible to observe more than about 12 dB of informational masking in that study. With threshold SN ratios in the range of −20 to −25 dB obtained here in the spatial conditions, there was much more room below the ceiling. It is important to recognize, however, that the threshold SN ratio is not the only relevant difference between the word detection and sentence recognition studies. The brevity of words relative to sentences may make them more difficult to perceptually extract from the background speech, thereby increasing informational masking in the nonspatial case. A more modest amount of informational masking was seen for the same two-talker masker using nonsense sentence detection (Helfer and Frey-

man, 2005), even though threshold SN ratios were also low in their spatial condition.

If one considers the detection threshold to be the lower bound of the intelligibility function and that detection of sentences would not be expected to be very different from detection of words taken from the sentences, then it must be noted that the intelligibility function takes quite an unusual form in the F-RF condition, requiring a sensation level of almost 30 dB before even 10% intelligibility could be achieved. This must be because, at threshold, the detection of speech in the F-RF condition is based on the awareness of any sound coming from the front loudspeaker; it does not need to be recognized as speech. Because of the impoverished spectral information in the six-channel vocoded nonsense sentences, the stimuli must apparently be well above threshold before any reasonable proportion of them can be understood. An analogous result was reported by Micheyl *et al.* (2006), who showed that the detection of a target complex against a competing complex occurred at a level nearly 20 dB below that required for accurate fundamental frequency discrimination. With the F-F condition, the detection threshold was much higher, so the span of levels between detection threshold and the beginnings of recognition was not so great.

One of the suppositions made in the introduction to these studies was that informational masking, as revealed by the amount of spatial release, might be increased with vocoded stimuli because the target-masker similarity and confusability would be increased. In experiment 2, spatial release from informational masking for vocoded stimuli was sizeable, but it was not larger than that observed for unprocessed speech (e.g., Fig. 2). This could be due to the fact that even the unprocessed two-talker female masker produced a great deal of informational masking, as revealed by the F-F versus F-RF difference of more than 20 dB. It is certainly possible that other speech maskers consisting of, for example, the speech of male talkers would produce considerably less informational masking in the unprocessed condition. Although not yet tested, it is hypothesized that vocoding would increase informational masking substantially in that situation.

D. Alternative interpretations

We believe that the above analysis explaining differences in results between the two experiments in terms of SN ratio offers the simplest account for the data and the one that is most consistent with the subjective impression of the authors when listening to the stimuli themselves. However, it must be recognized that achieving the desired difference of lower SN ratios in experiment 2 involved several other potentially important differences from experiment 1. In experiment 1, the target was a large set of sentences, none of which were repeated during a session, whereas experiment 2 used a smaller set of words excised from the sentences. In experiment 1 the task was recognition, whereas in experiment 2 it was detection. Finally, while the two-talker masker was identical across the experiments, in experiment 1 it was gated on and off with the target, whereas in experiment 2 it was presented continuously. These differences between experiments must be evaluated as potential alternative explanations for the differences in the results.

There is a substantial literature employing nonspeech stimuli that has demonstrated significant amounts of informational masking in a wide variety of tasks, including discrimination, identification, and detection (e.g., Kidd *et al.*, 1995, 1998; Oh and Lutfi, 1999; Oxenham *et al.*, 2003). Thus, there is nothing inherent to recognition tasks that should lead to the expectation of an absence of informational masking. The smaller stimulus set size of 20 excised words relative to the 320 sentences might have predicted less informational masking in the word detection experiment because of reduced target uncertainty. However, as noted above, our data from Balakrishnan and Freyman (2008) with unprocessed speech show that the 4AFC detection experiment with words does indeed give a larger spatial release from masking in the spatial masking conditions than we have observed with sentences in other studies (e.g., Freyman *et al.*, 2007). The main difference is a higher threshold SN ratio required in nonspatial detection of words relative to sentences. Most importantly, however, substantial and consistent improvements in the F-RF condition relative to the F-F condition have been observed with the very same target sentences, identical maskers, and masker gating characteristics, and the same recognition task used in the current study (Freyman *et al.*, 2007). Thus, when considered against all the literature that has preceded it, the finding in experiment 1 of no F-RF advantage must be due to some consequence of the vocoding process itself, as opposed to procedural differences between the two experiments reported in the current study.

The consequence of vocoding that has been emphasized in this paper was the higher SN ratio required for reasonable levels of speech recognition performance. However, there was also the possibility that the F-RF configuration did not produce the desired spatial perceptions with the vocoded stimuli that would be helpful in releasing informational masking. Informal listening described in Sec. II for experiment 1, as well as the results of experiment 2 showing large F-RF advantages with the identically processed vocoded stimuli, suggests that the spatial perceptions were available and could be useful. Finally, it is possible that the vocoding process eliminated informational masking in the recognition task, even though it apparently did not in the detection task, for reasons that have nothing to do with the SN ratios at which the stimuli were delivered. One possibility is that informational masking is reduced or eliminated because vocoding destroys the intelligibility of the two-talker masker. It is quite reasonable to consider that the intelligibility of the masker (and certainly that of the target) might be more important in recognition than in detection. Balakrishnan and Freyman (2008) showed that time-reversing a two-talker masker had almost no effect on detection in unprocessed speech conditions, whereas Freyman *et al.* (2001) showed that time-reversing a masker did improve recognition performance in the F-F condition. On the other hand, with both time-reversed maskers and foreign language maskers, a substantial spatial advantage in the F-RF condition remained. Thus, there is evidence of significant informational masking

occurring in recognition tasks in which the maskers were not intelligible to the listeners. Finally, it is also clear that the vocoding process does not by itself eliminate informational masking (e.g., Arbogast *et al.*, 2002, 2005).

In summary, while alternative explanations cannot be ruled out, to a large extent they require suppositions of interactions between variables that the literature provides little foundation for. On the other hand, explanations based on the notion of a ceiling for informational masking in the vicinity of 0 dB SN ratio are consistent with what appear to be truncations around 0 dB SN ratio in studies of both speech and nonspeech stimuli [e.g., Oxenham *et al.*, (2003) and Arbogast *et al.*, (2005), both discussed earlier in this paper]. The biggest challenge to our hypothesis is that there was no F-RF advantage even at +3 dB SN ratio, which is not very much above that hypothetical ceiling. However, it could simply be the case that +3 dB SN ratio (+6 dB above each individual talker in the two-talker complex) is sufficient to allow the target to stand out of the background. Further, because performance for the F-RF condition at +3 dB was less than 10% correct (Fig. 1), there was little room for the F-F condition to show worse performance. Finally, we would not want to argue with the notion that depending on the relevance of the information reaching the listener from a competing talker and message, softer interfering speech could sometimes be distracting, drawing attention away from a louder target talker (e.g., with a highly familiar competing talker or when one's name is spoken). However, this distraction may not be strongly related to similarity or uncertainty assumed to be involved in informational masking. Our interpretation is rather that louder target speech stands out from a background in a way that limits confusion between the target and a softer masker to a degree that makes it difficult to demonstrate further gains from the perception of target/masker spatial differences.

E. Implications

As noted in the Introduction, the type of noise-excited vocoded speech employed in the current studies has been used in the past to simulate important features of cochlear implant processing. Potential extensions of the results to wearers of those devices must be made with extreme caution, and there was no attempt to simulate the kind of spatial cues that implant users might receive. Nevertheless, the current experiments were interpreted to reveal a great deal of spatial release from informational masking with competing vocoded speech presented at poor SN ratios. Therefore, to the extent that spatial hearing could be partially restored in implant listeners through bilateral implantation, the present results suggest that there is the potential to provide this additional advantage beyond other benefits realized from bilateral implantation. At positive SN ratios, our interpretation of the absence of spatial release from masking is that there was no informational masking, even in nonspatial conditions. This, when considered in the context of other speech recognition studies that can be interpreted to show evidence of informational masking at positive SN ratios (Sec. IV B), presents an equivocal picture of informational masking and the role of spatial hearing at such SN ratios. It is likely that the role of informational masking is highly stimulus dependent. However, if our explanations for the current data are correct, this study adds evidence of target-to-masker intensity ratio dependence in addition to stimulus dependence. As implant processing improves and implant users have the expectation of succeeding in more difficult SN ratio conditions, the cautious prediction from the current study is that the challenge of informational masking will remain unless improvements in implant design include features that allow better talker recognition and segregation of voices.

ACKNOWLEDGMENTS

The authors would like to thank the Associate Editor, Andrew Oxenham, and two anonymous reviewers for their helpful comments on an earlier version of this manuscript. We also acknowledge J. Ackland Jones for his assistance with data collection and the support of the National Institute on Deafness and other Communication Disorders (DC 01625).

ANSI. (**1996**). *ANSI S3.6–1996; Specifications for Audiometers* (American National Standards Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Balakrishnan, U., and Freyman, R. L. (**2008**). "Speech detection in spatial and non-spatial speech maskers," J. Acoust. Soc. Am. **123**, 2680–2691.

Brungart, D. S., and Simpson, B. D. (**2002**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Kimberly, R. S. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meridith, R., Sirimanna, T., Tavarkiladze, G., Frolenkovgi, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Festen, M. 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.

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Freyman, R. L., Helfer, K. S., and Balakrishnan, U. (**2007**). "Variability and uncertainty in masking by competing speech," J. Acoust. Soc. Am. **121**, 1040–1046.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.

Gallun, F. J., Mason, C. R., and Kidd, G., Jr. (**2005**). "Binaural release from

informational masking in a speech identification task," J. Acoust. Soc. Am. **118**, 1614–1625.
George, E. L. J., Festen, J. M., and Houtgast, T. (**2006**). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 2295–2311.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.
Helfer, K. S. (**1997**). "Auditory and auditory-visual perception of clear and conversational speech," J. Speech Lang. Hear. Res. **40**, 432–443.
Helfer, K. S., and Freyman, R. L. (**2005**). "The role of visual speech cues in reducing energetic and informational masking," J. Acoust. Soc. Am. **117**, 842–849.
Kidd, G., Jr., Mason, C. R., and Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acta. Acust. Acust. **91**, 526–536.
Kidd, G., Jr. Mason, C. R., and Rohtla, T. L. (**1995**). "Binaural advantage for sound pattern identification," J. Acoust. Soc. Am. **98**, 1977–1986.
Kidd, G., Jr., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). "Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **104**, 422–431.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the information content of an irrelevant source diffeentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.
Micheyl, C., Bernstein, J. G. W., and Oxenham, A. J. (**2006**). "Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise," J. Acoust. Soc. Am. **120**, 1493–1505.
Nerbonne, G. P., Ivey, E. S., and Tolhurst, G. C. (**1983**). "Hearing protector evaluation in an audiometric testing room," Sound Vib. **17**, 20–22.
Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Oh, E. L., and Lutfi, R. A. (**1999**). "Informational masking by everyday sounds," J. Acoust. Soc. Am. **106**, 3521–3528.
Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G., Jr. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1453–1459.
Poissant, S. F., Whitmal, N. A., III., and Freyman, R. L. (**2006**). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," J. Acoust. Soc. Am. **119**, 1606–1615.
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.
Rakerd, B., Aaronson, N. L., and Hartmann, W. M. (**2006**). "Release from speech-on-speech masking by adding a delayed masker at a different location," J. Acoust. Soc. Am. **119**, 1597–1605.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shinn-Cunningham, B. G., Ihlefeld, A., Satyavarta, L., E. (**2005**). "Bottom-up and top-down influences on spatial unmasking," Acta. Acust. Acust. **91**, 967–979.
Stickney, G. S., Assman, P. F., Chang, J., and Zeng, F. G. (**2007**). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," J. Acoust. Soc. Am. **122**, 1069–1078.
Stickney, G. S., Zeng, F. G., Litovsky, R. Y., and Assman, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.

A sawtooth waveform inspired pitch estimator for speech and music

Arturo Camacho and John G. Harris
Computational NeuroEngineering Laboratory, University of Florida, Gainesville, Florida 32611

(Received 5 December 2007; revised 30 May 2008; accepted 2 June 2008)

A sawtooth waveform inspired pitch estimator (SWIPE) has been developed for speech and music. SWIPE estimates the pitch as the fundamental frequency of the sawtooth waveform whose spectrum best matches the spectrum of the input signal. The comparison of the spectra is done by computing a normalized inner product between the spectrum of the signal and a modified cosine. The size of the analysis window is chosen appropriately to make the width of the main lobes of the spectrum match the width of the positive lobes of the cosine. SWIPE′, a variation of SWIPE, utilizes only the first and prime harmonics of the signal, which significantly reduces subharmonic errors commonly found in other pitch estimation algorithms. The authors' tests indicate that SWIPE and SWIPE′ performed better on two spoken speech and one disordered voice database and one musical instrument database consisting of single notes performed at a variety of pitches.
 [DOI: 10.1121/1.2951592]

PACS number(s): 43.66.Hg [EJS] Pages: 1638–1652

I. INTRODUCTION

Pitch is an attribute of sound that gives important information about its source. In speech, it helps us to identify the gender of the speaker (females tend to have higher pitch than males) and gives additional meaning to words (a set of words may be interpreted as an affirmation or a question depending on the intonation). In music, it determines the names of the notes.

Several definitions of pitch have been proposed. One of them is "pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale" (American Standards Association, 1960). In practice, however, it is convenient not only to order sounds by their pitch, but also to assign a number to them. The standard method for doing that is to ask a group of listeners to adjust the frequency of a pure tone until its pitch matches the pitch of the target sound, and then define the pitch of the target sound as the frequency of the pure tone that best matched its pitch. Sometimes, complex tones made up of several harmonics are preferred to pure tones for matching.

Pitch estimation has applications in many areas that involve processing of sound. In music, it is used for automatic music transcription (Klapuri, 2004) and query by humming (e.g., Dannenberg *et al.*, 2004). In communications, it is used for speech coding (Spanias, 1994). In speech pathology, it is used to detect voice disorders (e.g., Yumoto *et al.* 1982). In linguistics, it is used to facilitate second language acquisition through the display of intonation patterns (de Bot, 1983).

Pitch estimation has a long history. An extensive review is given by Hess (1983). More recently, in a paper in which we presented a pitch estimator based on a smooth harmonic average peak-to-valley envelope (SHAPE) (Camacho and Harris, 2007), we included a review of some pitch estimators and illustrated some of their problems. Specifically, we showed that (i) algorithms that use the logarithm of the spectrum [e.g., cepstrum (Noll, 1967) and harmonic product spectrum (Schroeder, 1968)] are prone to fail when there are missing harmonics; (ii) algorithms that use the square of the spectrum [e.g., autocorrelation (Sondhi, 1968; Rabiner, 1977)] are prone to fail when there are salient harmonics; (iii) algorithms that give the same weight to all the harmonics (e.g., cepstrum, unbiased autocorrelation, and harmonic product spectrum) are prone to estimate the pitch as one of its subharmonics; (iv) algorithms that analyze the spectrum only at harmonic frequencies are prone to fail for inharmonic signals [e.g., harmonic product spectrum and subharmonic summation (Hermes, 1988)]; and (v) harmonic sieves (Duifhuis *et al.*, 1982), in which a component is accepted as harmonic if it is located within a certain range of a harmonic, are not completely satisfactory, since a slight shift of the component may put it in or out of the range, potentially changing the estimated pitch drastically.

The work presented here is an extension of SHAPE. The only new features are the use of window sizes that are proportional to the pitch period of the candidates, and the multiplication of the first and last negative lobes of the kernel by $1/2$ to avoid a bias that existed in SHAPE. It will be shown that the types of signals for which the algorithm is optimized are periodic signals whose spectral envelope decays inversely proportional to frequency. An example of such a signal is a sawtooth waveform. This type of signal is the one that motivated the name of the algorithm: sawtooth waveform inspired pitch estimator (SWIPE).

We conclude this section with a description of the scope of our work. Our goal is to estimate the pitch of speech and musical instruments, but not the pitch of some synthetic sounds that are traditionally used to test pitch perception models, such as sinusoidally amplitude-modulated noise (Meddis and Hewitt, 1991) and periodic signals with alternating phase harmonics (Meddis and O'Mard, 1997). Furthermore, we do not attempt to explain how the auditory system determines pitch but simply to create a black box that attempts to reproduce human pitch percepts. It is also worth

mentioning that our goal is to determine pitch but not fundamental frequency (defined as the maximum common divisor of its spectral components). In many cases, these two attributes coincide, but that is not always the case. For example, a periodic signal formed by the 13th, 25th, and 29th harmonics of 50 Hz (i.e., 650, 950, and 1250 Hz) is perceived as having a pitch of 334 or 650 Hz (Patel and Balaban, 2001) but not 50 Hz.

II. METHOD

A. Main idea

The main idea of the algorithm is the same underlying idea present in several pitch estimators (e.g., Sun, 2000; Rabiner, 1977; Sondhi, 1968; Noll, 1967): the measurement of the average peak-to-valley distance (APVD) at harmonic locations.[1] The APVD at the kth harmonic of f is defined as

$$\begin{aligned} d_k(f) &= \tfrac{1}{2}[|X(kf)| - |X((k-1/2)f)|] \\ &\quad + \tfrac{1}{2}[|X(kf)| - |X((k+1/2)f)|] \\ &= |X(kf)| - \tfrac{1}{2}[|X((k-1/2)f)| + |X((k+1/2)f)|], \end{aligned} \quad (1)$$

where $X(f)$ is the Fourier transform (FT) of the signal. Averaging over the first n peaks, the global APVD is

$$\begin{aligned} D_n(f) &= \frac{1}{n}\sum_{k=1}^{n} d_k(f) \\ &= \frac{1}{n}\left[\frac{1}{2}|X(f/2)| - \frac{1}{2}|X((n+1/2)f)| \right. \\ &\quad \left. + \sum_{k=1}^{n} |X(kf)| - |X((k-1/2)f)|\right]. \end{aligned} \quad (2)$$

As a first approach, we estimate the pitch as the frequency that maximizes the global APVD. This can be expressed using an integral transform as

$$p = \arg\max_f \int_0^\infty |X(f')| K_n(f,f') df', \quad (3)$$

where

$$\begin{aligned} K_n(f,f') &= \frac{1}{2}\delta(f' - f/2) - \frac{1}{2}\delta(f' - (n+1/2)f) \\ &\quad + \sum_{k=1}^{n} \delta(f' - kf) - \delta(f' - (k-1/2)f). \end{aligned} \quad (4)$$

Notice that the $1/n$ factor was obviated because the argument that maximizes the integral is invariant to scaling factors.

The kernel corresponding to a pitch candidate of 190 Hz and $n=9$ is shown in Fig. 1. The figure also shows the spectrum of a signal having the same pitch. The signal consists of the vowel /u/ passed through a bandpass filter that removed the frequencies outside the range 300–3400 Hz, mimicking telephone speech. Its spectrum exhibits a strong second harmonic that was presumably boosted by a formant close to 380 Hz.

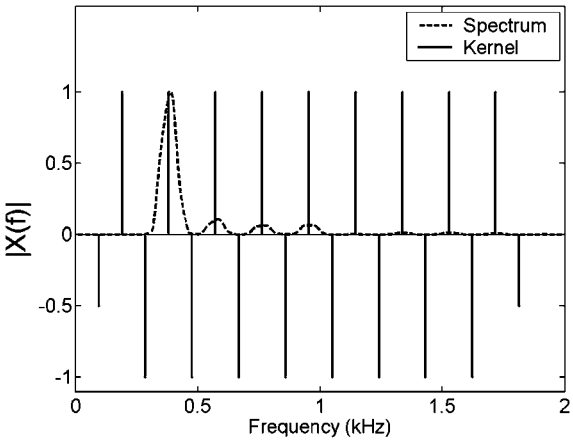


FIG. 1. Average peak-to-valley-distance (APVD) kernel. The APVD kernel has positive pulses at multiples of the fundamental and negative pulses in between. The first and last negative pulses have half the height of the others.

In the next sections, we will refine this first approach, trying to avoid the problem-causing features cited in the Introduction. Specifically, we will warp the spectrum, introduce a decaying weighting factor to the harmonics amplitudes, and replace the pulses with a smoother function.

B. Warping of the spectrum[2]

When we developed SHAPE, we found that using the square root of the spectrum produced better results than using the logarithm, square, or raw spectrum. Consequently, we will use it for SWIPE as well. The reason we believe the square root of the spectrum produces better results than the other functions will be postponed until the next section.

C. Weighting of the amplitude of the harmonics

Giving the same weight to all the harmonics amplitudes may lead to subharmonics of the pitch obtain the same score as the pitch. For example, if the signal consists of a pure tone of frequency f Hz and the same weight is applied to all the harmonic amplitudes, each of the subharmonic pitch candidates $f/2$, $f/3$,…, and f/n Hz, would have the same score as f Hz. To avoid this, we explored the use of the exponentially and harmonically decaying weights shown in Fig. 2. For exponential decays, a weight of r^{k-1} was applied to the kth harmonic amplitude ($k=1,2,\ldots,n$, and $r=0.5$, 0.7, and 0.9), and for harmonic decays, a weight of $1/k^p$ was applied to the kth harmonic amplitude ($k=1,2,\ldots,n$, and $p=1/2$, 1 and 2). In informal tests, the best results were obtained when using harmonic decays with $p=1/2$. Notice that this decay matches the decay of the square root of the average spectrum of vowels (Fant, 1970), which was the spectral warping shown to work best in the previous section. In other words, better pitch estimates were obtained when computing the inner product between the square root of the spectrum and a kernel whose envelope decays as $1/\sqrt{f}$, than when computing the inner products between the spectrum and a kernel whose envelope decays as $1/f$, for example.

The benefit of using a weighting of the harmonics amplitudes of the form $1/\sqrt{k}$ and the square root of the spectrum probably comes from the fact that when the signal has

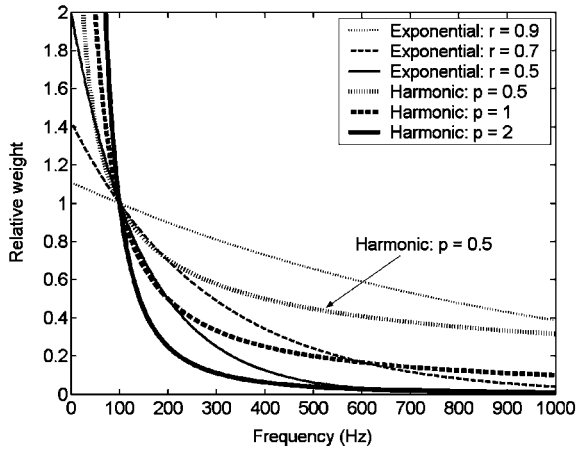


FIG. 2. Weighting of the harmonics amplitudes. Exponentially and harmonically decaying weighting factors of the form r^{k-1} (r=0.5, 0.7, and 0.9) and $1/k^p$ (p=1/2, 1 and 2) were utilized to weight the kth harmonic ($k=1,2,\ldots,n$). The highlighted curve corresponds to the one that produced the best results.

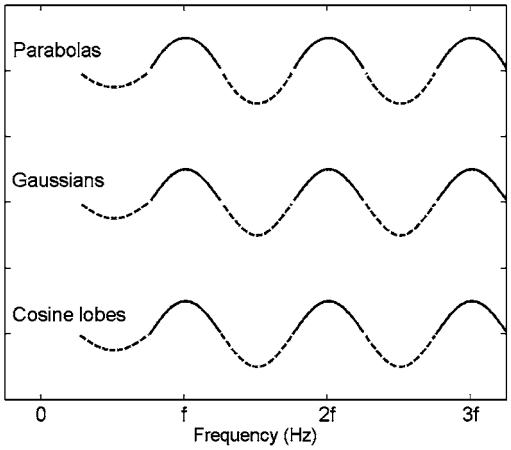


FIG. 3. Explored kernels. Kernels built from concatenations of truncated positive (continuous lines) and negative (dashed lines) parabolas, Gaussians, and cosine lobes were explored. (See the text for details about the selection of the truncation point.)

the expected shape of a vowel (i.e., the amplitudes of the harmonics decay as 1, 1/2, 1/3, etc.), each harmonic contributes to the inner product with a value proportional to its amplitude. If the spectrum has the aforementioned shape, the square root of the harmonics amplitudes will decay as 1, $1/\sqrt{2}$, and $1/\sqrt{3}$, etc., just like the weights of the harmonics amplitudes. This will make the terms in the sum of the inner product decay as 1, 1/2, 1/3, etc., and the relative contribution of each harmonic will be proportional to its amplitude. Conversely, if we compute the inner product over the raw spectrum using a weighting of the form $1/k$, the terms of the sum will be 1, 1/4, 1/9, etc., which are not proportional to the amplitude of the harmonics but to their square. This would make the contribution of the strongest harmonics too large and the contribution of the weakest too small.

D. Blurring of the harmonics

Analyzing the spectrum only at harmonic locations is inconvenient because it does not allow recognizing the pitch of inharmonic signals. To recognize the pitch of these signals, we propose the use of smooth weighting functions that take into account the spectrum not only at the harmonics, but also in their neighborhood.

In our previous paper (Camacho and Harris, 2007), it was shown that the local maxima of the kernel must be strictly concave (i.e., second derivative strictly positive). Consequently, concatenations of positive and negative truncated parabolas, Gaussians, and cosine lobes as shown in Fig. 3 (positive components with continuous lines and negative components with dashed lines) were proposed as kernels.

The criterion used to select the truncation point was the maximization of the smoothness of the concatenation by making as many derivatives continuous as possible. Even though smoothness sounds attractive, the main reason for using this criterion was the uniqueness of the solution for the Gaussian and the cosine: The truncation point has to be the inflection point. For the Gaussian, it occurs at one standard deviation and it warranties continuous first and second derivatives for the concatenation. For the cosine, it occurs at the zero crossing and it warranties all-order continuous derivatives (the concatenation of cosine lobes truncated in that way forms a cosine). In the case of the parabola, it is not possible to make a concatenation of positive and negative parabolas having continuous second and higher order derivatives, so arbitrarily we truncated it at its fixed point $(\pm 1,1)$.

As it can be suspected from the similarity of the kernels in Fig. 3, there were no significant differences in performance between them. However, we preferred the cosine kernel because of its smoothness and simplicity (it can be expressed without using concatenations).

E. Number of harmonics

With respect to the number of harmonics, there are typically two types of algorithms: those that use a fixed number of harmonics and those that use as many harmonics as possible, up to a certain frequency, usually the Nyquist frequency. We explored both possibilities and found that the best results were obtained when using as many harmonics as possible, although going beyond 3.3 kHz for speech and 5 kHz for musical instruments did not increase the performance significantly. Hence, we will use all available harmonics on SWIPE.

F. Warping of frequency

For the purpose of computing the integral of a function, we can think of warping the scale as the process of sampling the function more finely in some regions than others, effectively giving more emphasis to the more finely sampled regions. Since we are computing an inner product to estimate pitch, it makes sense to sample the spectrum more finely in the regions that contribute the most to the determination of pitch. It seems reasonable to assume that these regions are the ones with the most harmonic energy. In the case of speech, and assuming that the amplitude of the harmonics decays as $1/f$, it seems reasonable to sample the spectrum more finely in the neighborhood of the fundamental, and

decrease the granularity as we increase frequency, following the expected $1/f$ pattern. A decrease in granularity should also be performed below the fundamental because no harmonic energy is expected in that region. Unfortunately, the determination of the location of the fundamental is ill-posed, since that is precisely what we want to determine.

We explored frequency scale transformations of the form

$$\phi(f) = C \log(1 + f/\sigma), \tag{5}$$

where the constant factor C is irrelevant and can be set arbitrarily. The explored values of σ were 229, 700, and 0 (the latter in an asymptotical sense). These transformations correspond to the equivalent rectangular bandwidth (ERB) (Glasberg and Moore, 1990), mel, and logarithmic scales, respectively. (Notice also that the Hertz scale corresponds to $\sigma=\infty$.) We also explored the Bark scale given by the formula (Traunmüller, 1990)

$$z(f) = [26.81/(1 + 1960/f)] - 0.53. \tag{6}$$

To compute the inner product between the spectrum and the kernel, we sample both of them uniformly in the transformed scale before computing the inner product. The scale that on average produced the best results on speech was the ERB scale, which is expressed in a base-10 logarithmic scale as

$$\text{ERBs}(f) = 21.4 \log_{10}(1 + f/229). \tag{7}$$

This scale has several desirable characteristics: It approaches a logarithmic behavior as f increases, tends towards a constant (zero) as f decreases, and the frequency at which the transition occurs (229 Hz) is close to the mean fundamental frequency of speech, especially for females (Bagshaw, 1994; Wang and Lin, 2004; Schwartz and Purves, 2004).

For musical instruments, the Hertz scale was the one that produced the best results. An explanation of why this may be the case will be given in Sec. III D.

G. Window type and size

Most common methods of pitch estimation use a fixed window size. This makes the width of the main lobe of each of the spectral components to be fixed as well. This has the disadvantage of making high pitches more likely to obtain high scores than low pitches. The reason for this is illustrated in Fig. 4. Figure 4(a) shows the spectrum of a signal with a pitch of 500 Hz and the kernel corresponding to that pitch, and Fig. 4(b) shows the spectrum of a signal with a pitch of 125 Hz and the kernel corresponding to that pitch. The width of the cosine lobes increases with pitch, but not the width of the spectral lobes. This causes the overall weight given to the main spectral lobes to increase as the pitch increases, and to decrease as the pitch decreases. Actually, below a certain pitch (125 Hz in this example), the weight given to the sides of the main spectral lobes becomes negative (not shown in the figure).

One way to solve this problem is to make the spectral lobes as narrow as possible, but this requires making the window infinitely large, which is undesirable given the changing nature of pitch (signals with constant pitch for eternity do not occur in real life). Another way to solve the problem is to try to make the square root of the main spectral lobes match the cosine lobes for each of the pitch candidates, and then compute the normalized inner product between them, defined as

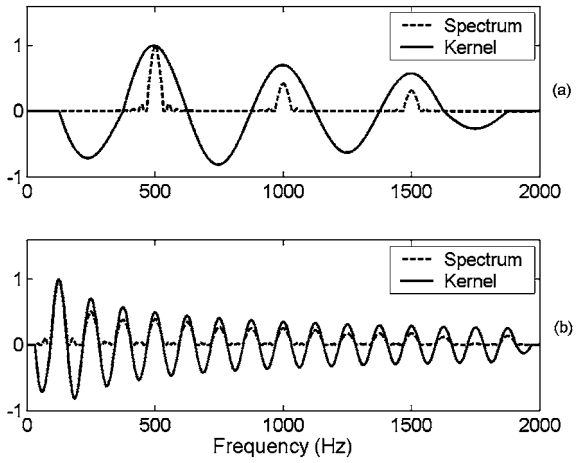


FIG. 4. Similarity between cosine lobes and square root of main spectral lobes. (a) High pitch, low similarity. (b) Low pitch, high similarity.

$$\frac{\int_R \xi(f)\psi(f)df}{(\int_R \xi^2(f)df)^{1/2}(\int_R \psi^2(f)df)^{1/2}}, \tag{8}$$

for any two functions $\xi(f)$ and $\psi(f)$, over a region R. Since the spectrum is non-negative, but not the cosine, the cosine kernel must be normalized using only its positive part in order to obtain a normalized inner product close to 1.[3]

A type of window whose square-root spectrum has a large similarity with a cosine is the Hann window. A Hann window of size T (in seconds) is defined as

$$h_T(t) = \frac{1}{T}\left[1 + \cos\left(\frac{2\pi t}{T}\right)\right] \tag{9}$$

for $|t| < 1/2$, and 0 otherwise. Its FT is

$$H_T(f) = \text{sinc}(Tf) + \tfrac{1}{2}\text{sinc}(Tf - 1) + \tfrac{1}{2}\text{sinc}(Tf + 1), \tag{10}$$

where the sinc function is defined as

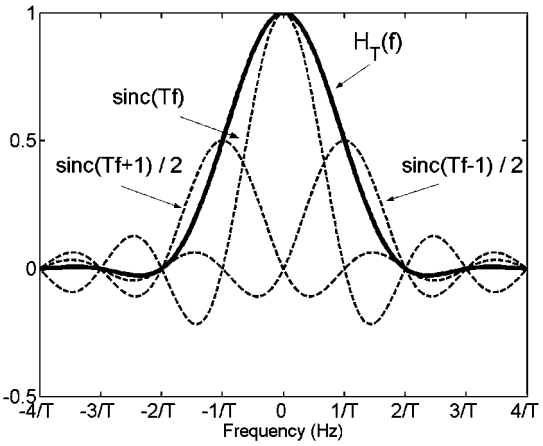


FIG. 5. FT of the Hann window. The FT of the Hann window is a sum of three sinc functions.

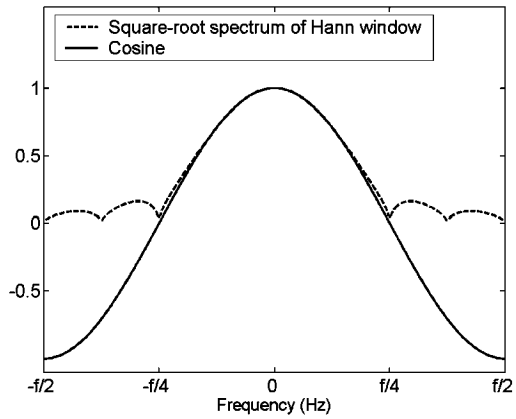


FIG. 6. Similarity between cosine lobe and square root of the spectrum of the Hann window.

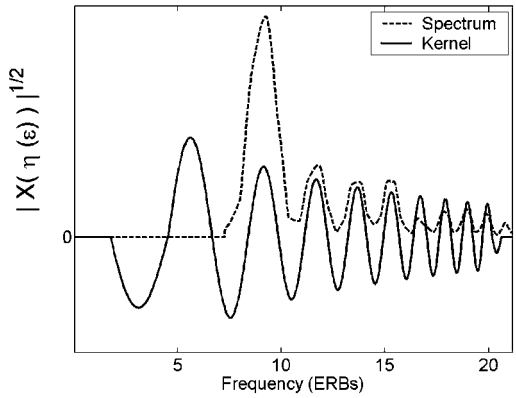


FIG. 7. Normalized SWIPE kernel. The SWIPE kernel consists of a truncated decaying cosine with halved first and last valleys. The kernel is normalized using its positive part.

$$\mathrm{sinc}(\varphi) = \frac{\sin(\pi\varphi)}{\pi\varphi}. \tag{11}$$

This FT is illustrated in Fig. 5. The width of its main lobe is $4/T$. If we match this width to the width of the cosine lobe, $f/2$, where f is the pitch of the candidate (in hertz), and solve for T, we will find that the window size must be $T=8/f$.

Figure 6 shows the square root of the spectrum of a Hann window of size $T=8/f$ and a cosine with period f. The similarity between the main lobe of the spectrum and the positive lobe of the cosine is remarkable. They match at five frequencies: 0, $\pm f/8$, and $\pm f/4$, with values $\cos(0)=1$, $\cos(\pi/4)=1/\sqrt{2}$, and $\cos(\pi/2)=0$, respectively. The normalized inner product between the main lobe of the spectrum and the positive part of the cosine sampled at 128 equidistant points is 0.9996, and the normalized inner product computed over the whole period of the cosine sampled at 256 equidistant points is 0.8896. This reduction in normalized inner product is caused by the side lobes.

The normalized inner product between the cosine lobe and the main lobe of the square root of the spectrum for other types of windows is shown in Table I. The width of the main lobes of the spectrum of these windows is $2k/T$, where k depends on the window type (see Oppenheim *et al.*, 1999), and is given in the second column of the table.

The window type that produces the largest normalized inner product between the square root of the main lobe of the spectrum and the cosine lobe is the Hann window (0.9996). When computed over the whole period of the cosine, there are other window types that produce a larger normalized inner product. However, most of these windows are larger than the Hann window, and therefore require more computation. The only window that has the same value of k as the Hann window but a larger normalized inner product over the whole period of the cosine is the Hamming window. However, we prefer the Hann window because of its simpler formula.

H. SWIPE

Putting all pieces together, we can express the SWIPE estimate of the pitch at time t as

$$p(t) = \arg\max_{f} \frac{\int_0^{\mathrm{ERBs}(f_{\max})} 1/\eta(\varepsilon)^{1/2} K(f,\eta(\varepsilon)) |X(t,f,\eta(\varepsilon))|^{1/2} d\varepsilon}{\left(\int_0^{\mathrm{ERBs}(f_{\max})} 1/\eta(\varepsilon)[K^{+}(f,\eta(\varepsilon))]^2 d\varepsilon\right)^{1/2} \left(\int_0^{\mathrm{ERBs}(f_{\max})} |X(t,f,\eta(\varepsilon))| d\varepsilon\right)^{1/2}}, \tag{12}$$

where

$$K(f,f') = \begin{cases} \cos(2\pi f'/f) & \text{if } 3/4 < f'/f < n(f) + 1/4 \\ \frac{1}{2}\cos(2\pi f'/f) & \text{if } 1/4 < f'/f < 3/4 \text{ or } n(f) + 1/4 < f'/f < n(f) + 3/4 \\ 0 \quad \text{otherwise,} & \end{cases} \tag{13}$$

$$X(t,f,f') = \int_{t-4/f}^{t+4/f} [1 + \cos(\pi f(t'-t)/4)] x(t') e^{-j2\pi f' t'} dt', \tag{14}$$

ε is frequency in ERBs, $\eta(\cdot)$ converts frequency from ERBs into hertz, ERBs$(\cdot)$ converts frequency from hertz into ERBs, $K^{+}(\cdot)$ is the positive part of $K(\cdot)$ {i.e., $\max[0,K(\cdot)]$}, $f_{\max}$ is the maximum frequency to be used (typically the Nyquist frequency, although 5 kHz is enough for most applications), $n(f)=\lfloor f_{\max}/f-3/4 \rfloor$, and $j^2=-1$.

The normalized kernel corresponding to a candidate with frequency 190 Hz (5.6 ERBs) is shown in Fig. 7. The

TABLE I. Normalized inner products between the kernel and the square root of the spectrum of several window types,[a] computed over the main spectral lobe and over one period of the cosine around zero. The parameter k determines the window size T required to produce those normalized inner products, based on the formula $T=4k/f$, where f is the frequency of the cosine.

		Normalized inner product	
Window type	k	Main spectral lobe	Whole cosine period
Bartlett	2	0.9984	0.7959
Bartlett–Hann	2	0.9995	0.8820
Blackman	3	0.9899	0.9570
Blackman–Harris	4	0.9738	0.9689
Bohman	3	0.9926	0.9474
Flattop	5	0.9896	0.9726
Gauss	3.14	0.9633	0.8744
Hamming	2	0.9993	0.9265
Hann	2	0.9996	0.8896
Nuttall	4	0.9718	0.9682
Parzen	4	0.9627	0.9257
Rectangular	1	0.9925	0.5236
Triangular	2	0.9980	0.8820

[a]The normalized inner products were computed using 128 equidistant samples for the main spectral lobe and 256 equidistant samples for the whole cosine period.

figure also shows the normalized spectrum of a signal with the same pitch. It is easy to show that the type of signal for which the function maximized in Eq. (12) achieves its maximum is periodic signals whose spectral envelope decays as $1/f$. An example of such type of signal is a sawtooth waveform, which is the one that inspired the name of the algorithm. Another type of signal with that property (on average) is a vowel (Fant, 1970).

I. SWIPE′

One of the most common mistakes of SWIPE (and other algorithms) is misestimating the pitch as one of its subharmonics. Figure 8 illustrates why this error is common. It shows the spectrum of a signal whose component frequencies are harmonics of 100 Hz and are of equal amplitude. It

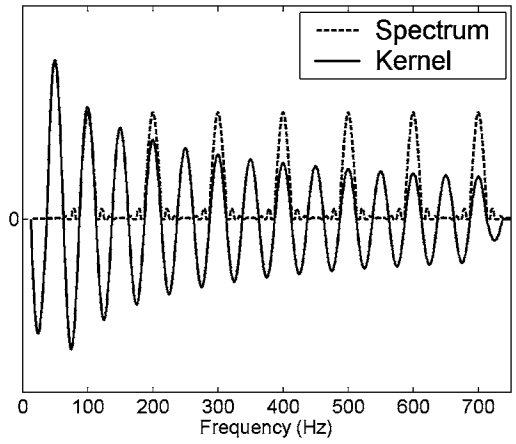


FIG. 8. Kernel corresponding to a subharmonic of the pitch. The figure shows a signal formed by equal-amplitude harmonics of 100 Hz, and the kernel corresponding to the 50 Hz candidate. This kernel produces a high score based only on its even components.

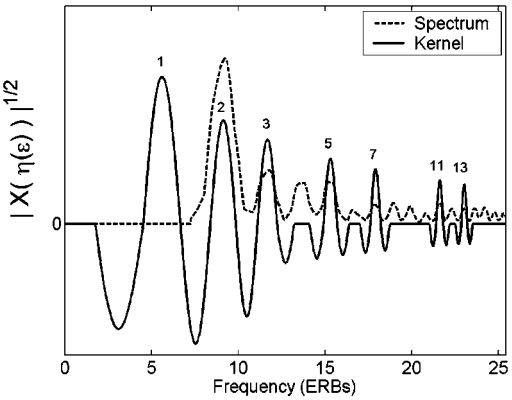


FIG. 9. Normalized SWIPE′ kernel. The SWIPE′ kernel has peaks only on its first and prime harmonics.

also shows the kernel corresponding to a candidate of 50 Hz. Since every multiple of 100 is also a multiple of 50, this candidate will receive significant credit based only on its even harmonics. Something similar would occur with the kernel at 33 Hz (not shown), but based on its multiples-of-3 harmonics. The phenomenon extends to any subharmonic of the pitch.

To avoid such situations, we propose to remove from the kernel its nonprime harmonics, except the first one. We do this by redefining the kernel as

$$K(f,f') = \sum_{i\in\{1\}\cup P} K_i(f,f'), \tag{15}$$

where P is the set of prime numbers, and

$$K_i(f,f') = \begin{cases} \cos(2\pi f'/f) & \text{if } |f'/f - i| < 1/4 \\ \frac{1}{2}\cos(2\pi f'/f) & \text{if } 1/4 < |f'/f - i| < 3/4 \\ 0 \quad \text{otherwise.} \end{cases} \tag{16}$$

[Notice that the SWIPE kernel can also be written as in Eqs. (15) and (16) by including all the harmonics (not only the first and the primes) in the sum of Eq. (15).]

This variation of SWIPE in which only the first and prime harmonics are used is named SWIPE′. The kernel corresponding to a pitch candidate of 190 Hz (5.6 ERBs) is shown in Fig. 9. The numbers on top of the peaks of the kernel indicate the corresponding harmonic number. With this approach, no candidate below the pitch will get credit from more than one of the harmonics of the signal.

J. Reduction of computational cost

Equations (12)–(14) [or Eqs. (12) and (14)–(16)] form a complete description of the algorithm to estimate pitch in continuous time. However, in practice, we use digital computers, which require discrete-time and discrete-frequency versions of these equations. Also, computational power is usually limited, which implies that arbitrarily large levels of resolution are unattainable. However, most signals we deal with are relatively slowly varying, which means that after a certain point, increments in resolution are unnecessary because they only provide redundant information. For many

applications, computations can be done at a relatively low resolution, and if higher resolution is required, interpolation can be used to fill in the gaps. This idea will be exploited in the following sections.

1. Reduction of the number of Fourier transforms

One of the most costly operations of SWIPE and SWIPE′ is the computation of the (short term) FT in Eq. (14). Hence, to reduce computational cost, it is important to reduce the number of FTs. There are two strategies to achieve this: reducing the window overlap and sharing FTs among candidates.

a. Reduction of window overlap. The most common windows used in signal processing are attenuated towards zero at the edges. Therefore, it is possible to overlook short events if they are located at these edges. To avoid this situation, it is common to use overlapping windows, which increase the coverage of the signal at the cost of an increase in computation. However, after a certain point, overlapping windows produce redundant data, making it unnecessary to go to the limit of sample by sample window shifting.

A study by Doughty and Garner (1947) showed that, depending on frequency, at least two to four cycles are required to perceive the pitch of a pure tone. To avoid this interaction between number of cycles and frequency, we will assume that a minimum of four cycles are required to detect the pitch of a pure tone, regardless of its frequency. Furthermore, we extrapolate this result to the type of signals for which SWIPE produces maximum output: sawtooth waveforms. Finally, we assume that the pitched/unpitched decision threshold is no larger than half the score obtained by a full length sawtooth waveform (~0.89, according to Table I). It can be shown that, under these assumptions, four cycles of a sawtooth waveform will obtain a score higher than the threshold as long as the window overlap is at least 50%.

The worst case scenario occurs when the four cycles of the sawtooth are centered at the middle point between two consecutive windows. In this scenario, the signal spans over one-half of each of the windows and the score obtained on each window is half the score that would be obtained if the signal would cover the whole window. Shifting the signal to the left/right necessarily increases the score on the left/right window, making the detection of the pitch more likely.

b. Use of exclusively power-of-2 window sizes. The use of the "optimal" window size proposed in Sec. II G requires that each pitch candidate uses a different window size, which means that a FT must be computed for each candidate. Furthermore, since the most efficient way to compute a FT is using a fast Fourier transform (FFT) and this algorithm is more efficient when using window sizes that are powers of 2 (in samples), the optimal window sizes may not be appropriate to use a FFT.

To alleviate this problem, we propose to substitute the optimal window size with the two closest power-of-2 window sizes, and then linearly combine the scores obtained using these windows into a single score, based on the distance between the size of the windows and the optimal window size, in a logarithmic scale. More precisely, if the optimal window size in samples is $N^*=2^{L+\lambda}$, where L is an integer and $0\leq\lambda<1$, the two closest power-of-2 window sizes are $N_0=2^L$ and $N_1=2^{L+1}$. Using these sizes, we compute scores $S_0(f)$ and $S_1(f)$ as in the function maximized in Eq. (12), modifying the integration region in Eq. (14) to have a length corresponding to those sizes. Finally, we combine these two scores into the single score

$$S(f)=(1-\lambda)S_0(f)+\lambda S_1(f). \tag{17}$$

The intuition behind this formula is that if the optimal window size is closer to N_0 than to N_1, $S_0(f)$ will have a larger contribution towards $S(f)$ than $S_1(f)$, and vice versa.

Alternatively, we could have used only the score produced by the closest power-of-2 window size (in a logarithmic scale), but this would have introduced undesirable discontinuities in $S(f)$, as shown in Fig. 10(a). This figure shows scores produced by two different window sizes. The pitch was chosen to match the point where the change of

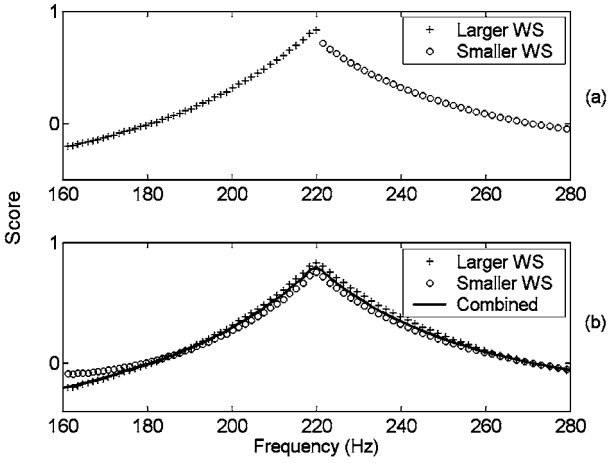


FIG. 10. Individual and combined scores of candidates between 160 and 280 Hz for a sawtooth waveform with a pitch of 220 Hz. The scores in (a) were produced using the power-of-2 window size closest to the optimal window size for each of the candidates. The scores in (b) (crosses and circles) were produced using the *two* power-of-2 window sizes closest to the optimal window size for each of the candidates. The continuous line in (b) consists of the combination of the scores produced by the two power-of-2 window sizes closest to the optimal window size for each of the candidates, as determined in Eq. (17).

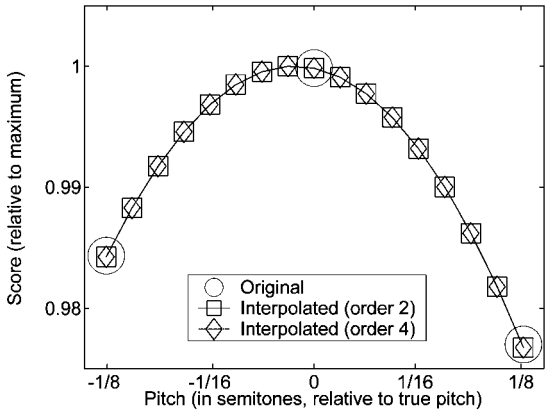


FIG. 11. Interpolated scores. Circles show scores of candidates separated at a distance of 1/8 semitone, squares show interpolated scores produced by an order-2 polynomial, and diamonds show interpolated scores produced by an order-4 polynomial.

window size occurs. Since the larger window tends to produce larger scores, the estimated pitch may be shifted slightly to the left in this case. In Fig. 10(b), the scores produced by both windows were combined using Eq. (17) to produce the continuous solid curve in between.

Besides using a convenient window size for the FFT computation, the approximation of the optimal window size using the two closest power-of-2 window sizes has another advantage that is probably more important: The same FFT can be shared by several pitch candidates, more precisely, by all the candidates within an octave of the optimal pitch for that FFT.

Using this approach, and translating the algorithm to a discrete-time domain (necessary to compute a FFT), we can express the SWIPE estimate of the pitch at the discrete-time index τ as

$$p[\tau] = \arg\max_f (1 - \lambda(f)) S_{L(f)}(\tau, f) + \lambda(f) S_{L(f)+1}(\tau, f), \tag{18}$$

where

$$\lambda(f) = L^*(f) - L(f), \tag{19}$$

$$L(f) = \lfloor L^*(f) \rfloor, \tag{20}$$

$$L^*(f) = \log_2(8 f_s / f), \tag{21}$$

$$S_L(\tau, f) = \frac{\sum_{m=0}^{\lfloor \mathrm{ERBs}(f_{\max})/\Delta\varepsilon \rfloor} 1/\eta(m\Delta\varepsilon)^{1/2} K(f, \eta(m\Delta\varepsilon)) |\hat{X}_{2^L}[\tau, \eta(m\Delta\varepsilon)]|^{1/2}}{\left(\sum_{m=0}^{\lfloor \mathrm{ERBs}(f_{\max})/\Delta\varepsilon \rfloor} 1/\eta(m\Delta\varepsilon) [K^+(f, \eta(m\Delta\varepsilon))]^2\right)^{1/2} \left(\sum_{m=0}^{\lfloor \mathrm{ERBs}(f_{\max})/\Delta\varepsilon \rfloor} |\hat{X}_{2^L}[\tau, \eta(m\Delta\varepsilon)]|\right)^{1/2}}, \tag{22}$$

$$\hat{X}_N[\tau, f'] = I(\{0, \ldots, N-1\}, X_N[\tau, \{0, \ldots, N-1\}], f' N / f_s), \tag{23}$$

$$X_N[\tau, \varphi] = \sum_{\tau'=-N/2}^{N/2-1} [1 + \cos(2\pi(\tau' - \tau)/N)] x[\tau'] e^{-j 2\pi \varphi \tau' / N}, \tag{24}$$

f_s is the sampling frequency, $\Delta\varepsilon$ is the ERB scale step size (0.1 was used in our tests), and $I(\Phi, \Xi, \phi)$ is an interpolating function that uses the functional relations $\Xi_k = F(\Phi_k)$ to predict the value of $F(\phi)$. The other variables, constants, and functions are defined as before (see Sec. II H).

2. Reducing the number of spectral integral transforms

Since the pitch resolution of SWIPE depends on the granularity of the pitch candidates, to achieve high pitch resolution, a large number of pitch candidates are required. To avoid excessive computational cost, we propose to compute the score only for certain candidates, and then use interpolation to estimate the score of the others. Since the autocorrelation function of a signal is the FT of its power spectrum, it consists of a sum of cosines that can be approximated around zero by using a Taylor series expansion with even powers (de Cheveigné, 2002). If the signal is periodic, its autocorrelation function is also periodic, and the shape of the curve around the pitch period is the same as the shape around zero. This means that the autocorrelation function around the pitch period can be approximated by the Taylor series expansion around zero after shifting it to the pitch period. If the width of the spectral lobes is narrow and the energy of the high frequency components is small, the terms of order 4 and higher in the series vanish as the independent variable approaches the pitch period, and the series can be approximated using a parabola.

Since SWIPE multiplies a compressed version of the spectrum by a cosine-based kernel, it could be expected that a similar argument applies to SWIPE as well. However, there are two complications: First, the widths of the spectral lobes produced by SWIPE are not narrow, in fact, they are as wide as the positive lobes of the cosine; and second, the use of the square root of the spectrum rather than its energy could make the contribution of the high frequency components large, violating the principle of low contribution of high frequency components.

Nevertheless, parabolic interpolation produces a good fit to SWIPE's score in the neighborhood of the maxima, as shown in Fig. 11. This figure shows in circles the scores (relative to the maximum) corresponding to the pitch and two pitch candidates located at a distance of 1/8 semitones from the pitch, for a signal consisting of a sawtooth waveform. Since the function looks almost symmetric around its maximum, its Taylor series expansion must contain mostly even terms. The figure also shows polynomial expansions of order 2 (squares) and order 4 (diamonds), which look identical. Expansions of higher order are not shown, but they look the same as these two. Therefore, a parabola is good enough to satisfactorily interpolate the scores in the neighborhood of the maximum for the chosen resolution.

III. EVALUATION

SWIPE and SWIPE′ were compared against other algorithms in terms of performance using three speech databases and a musical instruments database.

A. Algorithms

The algorithms against which SWIPE and SWIPE′ were compared are the following:

AC-P. This algorithm (Boersma, 1993) computes the autocorrelation of the signal and divides it by the autocorrelation of the window used to analyze the signal. It uses postprocessing to reduce discontinuities in the pitch trace. It is available with the Praat System at http://www.fon.hum.uva.nl/praat. The name of the function is *ac*.

AC-S. This algorithm uses the autocorrelation of the cubed signal. It is available with the Speech Filing System at http://www.phon.ucl.ac.uk/resource/sfs. The name of the function is *fxac*.

ANAL. This algorithm (Secrest and Doddington, 1983) uses autocorrelation to estimate the pitch, and dynamic programming to remove discontinuities in the pitch trace. It is available with the Speech Filing System at http://www.phon.ucl.ac.uk/resource/sfs. The name of the function is *fxanal*.

CATE. This algorithm uses a quasiautocorrelation function of the speech excitation signal to estimate the pitch. We implemented it based on its original description (Di Martino and Laprie, 1999). The dynamic programming component used to remove discontinuities in the pitch trace was not implemented.

CC. This algorithm uses cross correlation to estimate the pitch and postprocessing to remove discontinuities in the pitch trace. It is available with the Praat System at http://www.fon.hum.uva.nl/praat. The name of the function is *cc*.

CEP. This algorithm (Noll, 1967) uses the *cepstrum* of the signal and is available with the Speech Filing System at http://www.phon.ucl.ac.uk/resource/sfs. The name of the function is *fxcep*.

ESRPD. This algorithm (Bagshaw, 1993; Medan *et al.*, 1991) uses a normalized cross correlation to estimate the pitch, and postprocessing to remove discontinuities in the pitch trace. It is available with the Festival Speech Filing System at http://www.cstr.ed.ac.uk/projects/festival. The name of the function is *pda*.

RAPT. This algorithm (Secrest and Doddington, 1983) uses a normalized cross correlation to estimate the pitch, and dynamic programming to remove discontinuities in the pitch trace. It is available with the Speech Filing System at http://www.phon.ucl.ac.uk/resource/sfs. The name of the function is *fxrapt*.

SHS. This algorithm (Hermes, 1988) uses subharmonic summation. It is available with the Praat System at http://www.fon.hum.uva.nl/praat. The name of the function is *shs*.

SHR. This algorithm (Sun, 2000) uses the subharmonic-to-harmonic ratio. It is available at Matlab Central (http://www.mathworks.com/matlabcentral) under the title "Pitch Determination Algorithm." The name of the function is *shrp*.

TEMPO. This algorithm (Kawahara *et al.*, 1999) uses the instantaneous frequency of the outputs of a filterbank. It is available with the STRAIGHT System at its author web page (http://www.wakayama-u.ac.jp/~kawahara). The name of the function is *exstraightsource*.

YIN. This algorithm (de Cheveigné and Kawahara, 2002) uses a modified version of the average squared difference function. It was made available by his author. The name of the function is *yin*.

B. Databases

The databases used to test the algorithms were the following.

DVD: disordered voice database. This database contains 657 samples of the sustained vowel "ah" produced by persons with disordered voice. Most of the files have a sampling rate of 25 kHz, although a few of them (77) have a sampling rate of 50 kHz. It can be bought from Kay Pentax (http://www.kayelemetrics.com).

KPD: Keele pitch database. This speech database was collected by Plante *et al.* (1995) at Keele University with the purpose of evaluating pitch estimation algorithms. It contains about 8 min of speech spoken by five males and five females, sampled at 20 kHz. Laryngograph data were recorded simultaneously with speech and were used to produce estimates of the fundamental frequency.

MIS: musical instruments samples. This database contains more than 150 min of sound produced by 20 different musical instruments sampled at 44.1 kHz.[4] It was collected at the University of Iowa Electronic Music Studios, directed by Lawrence Fritts, and is publicly available at http://theremin.music.uiowa.edu.

PBD: Paul Bagshaw's database for evaluating pitch determination algorithms. This database contains about 8 min of speech spoken by one male and one female, sampled at 20 kHz. Laryngograph data were recorded simultaneously with speech and were used to produce estimates of the fundamental frequency. It was collected by Paul Bagshaw at the University of Edinburg (Bagshaw *et al.*, 1993; Bagshaw, 1994) and is publicly available at http://www.cstr.ed.ac.uk/research/projects/fda.

C. Methodology

Whenever possible, the algorithms were asked to produce pitch estimates at every millisecond. The search range was set to 40–800 Hz for speech and 30–1666 Hz for musical instruments, and the algorithms were given the freedom to decide if the sound was pitched or not. To compute the statistics, we rounded to the closest millisecond the times associated with each ground truth value and the times associated with each pitch estimate produced by the algorithms that were able to give an output every millisecond, and considered only the rounded times at which all these algorithms and the ground truth agreed that the sound was pitched.

Special care was taken to account for misalignments. Specifically, the pitch estimates were associated with the time corresponding to the center of their respective analysis windows, and when the ground truth pitch varied over time (i.e., for PBD and KPD), the estimated pitch time series were shifted in steps of 1 ms within the range ± 100 ms to find the best alignment with the ground truth.

The performance measure used to compare the algorithms was the gross error rate (GER). A gross error occurs when the estimated pitch is off from the reference pitch by

TABLE II. GERs obtained for speech and voice databases (PBD, KPD, and DVD).[a]

Algorithm	Gross error (%)			
	PBD	KPD	DVD	Average
SWIPE′	0.13	0.83	0.63	0.53
SHS	0.15	1.00	1.10	0.75
SWIPE	0.15	0.87	1.70	0.91
RAPT	0.75	1.00	2.40	1.40
TEMPO	0.32	1.90	2.00	1.40
YIN	0.33	1.40	4.50	2.10
SHR	0.69	1.50	5.10	3.50
ESRPD	1.40	3.90	4.60	5.00
CEP	6.10	4.20	14.00	5.90
AC-P	0.73	2.90	16.00	6.70
CATE	2.60	10.00	7.20	6.60
CC	0.48	3.60	5.00	2.40
ANAL	0.83	2.00	35.00	13.00
AC-S	8.80	7.00	40.00	19.00
Average	1.70	3.00	9.90	4.90

[a]Values computed using two significant digits.

TABLE III. GERs by gender for speech databases (PVD and KPD).[a]

Algorithm	Gross error (%)		
	Male	Female	Average
SWIPE′	0.36	2.40	1.4
SHS	0.55	2.50	1.5
SWIPE	0.49	2.70	1.6
RAPT	0.42	2.90	1.7
TEMPO	0.67	3.10	1.9
SHR	0.61	3.60	2.1
YIN	1.10	3.20	2.2
AC-P	2.10	3.60	2.9
CEP	1.80	4.20	3.0
CC	2.40	4.50	3.5
ESRPD	3.10	3.90	3.5
ANAL	1.30	5.90	3.6
AC-S	3.20	10.00	6.6
CATE	11.00	4.20	7.6
Average	2.10	4.00	3.1

[a]Values computed using two significant digits.

more than 20%. At first glance, this margin of error may seem too large, but considering that most of the error pitch estimation algorithms produced are octave errors (i.e., halving or doubling the pitch), this is a reasonable metric. On the other hand, this tolerance gives room for dealing with misalignments. The GER measure has been used previously to test pitch estimators by other researchers (e.g., Bagshaw *et al.*, 1993; Di Martino and Laprie, 1999; de Cheveigné and Kawahara, 2002).

D. Results

Table II shows the GERs for each of the algorithms over each of the speech databases. The rows and the columns are sorted by average GER (the best algorithms are at the top and the more difficult databases are at the right). The best algorithm overall is SWIPE′, followed by SHS and SWIPE. On average, SHS performed better than SWIPE; however, SHS beat SWIPE only on the disordered voice database, but not in the normal speech databases, which suggests that SWIPE works better than SHS on normal speech.

Table III shows the pitch estimation performance as a function of gender for the two databases for which we had access to this information: PVD and KPD. On average, error rates are larger for females than for males.

Table IV shows the GERs for the musical instruments database. Some of the algorithms were not evaluated on this database because they did not provide a mechanism to set the search range, and the range they covered was smaller that the pitch range spanned by the database. The two algorithms that performed the best were SWIPE′ and SWIPE.

Table V shows the GERs by instrument family. The two best algorithms are SWIPE′ and SWIPE. SWIPE′ tends to perform better than SWIPE, except for the piano, for which SWIPE produces almost no error. SWIPE′ performance on piano is relatively bad, especially when compared against correlation based algorithms. In general, the family for which fewer errors were obtained was the brass family (many algorithms achieved almost perfect performance for this family). The family for which more errors were produced was the strings family playing *pizzicato*, i.e., by plucking the strings. Indeed, pizzicato sounds were the ones for which the performers produced more errors and the ones that were hardest for us to label (see Appendix).

Table VI shows the GERs as a function of octave. The best performance on average was achieved by SWIPE′ and SWIPE.

As a final test, we wanted to validate the choices we made in Sec. II, i.e., shape of the kernel, warping of the spectrum, weighting of the harmonics, warping of the frequency scale, and selection of window type and size. For this purpose, we evaluated SWIPE′ replacing every one of its features with most[5] of the alternative features described in Secs. II B–II G. We varied each of these features, one at a time, and obtained the results shown in Table VII. The step sizes used for each of the alternative frequency scales and the actual number of steps is shown in Table VIII. The step sizes

TABLE IV. GERs for musical instruments (MIS database).[a]

Algorithm	Gross error (%)		
	Underestimates	Overestimates	Total
SWIPE′	1.00	0.10	1.10
SWIPE	1.30	0.02	1.30
SHS	0.88	1.00	1.90
TEMPO	0.29	1.70	2.00
YIN	1.60	0.83	2.40
AC-P	3.20	0.00	3.20
CC	3.60	0.00	3.60
ESRPD	5.30	1.50	6.80
SHR	15.00	5.30	20.00
Average	3.60	1.20	4.70

[a]Values computed using two significant digits.

TABLE V. GERs by instrument family (MIS database).[a]

	Gross error (%)					
Algorithm	Brass[b]	Bowed strings[c]	Woodwinds[d]	Piano	Plucked strings[e]	Average
SWIPE′	0.01	0.19	0.14	2.20	8.80	2.30
SWIPE	0.00	0.22	0.23	0.02	11.00	2.30
TEMPO	0.00	2.60	1.40	7.30	4.00	3.10
YIN	0.03	1.10	1.50	0.36	14.00	3.40
SHS	0.02	1.50	0.72	12.00	8.10	4.50
AC-P	0.03	0.56	0.80	0.36	26.00	5.60
CC	0.07	0.83	1.00	0.36	28.00	6.00
ESRPD	4.00	6.90	7.10	6.00	11.00	7.00
SHR	22.00	25.00	38.00	26.00	15.00	25.00
Average	2.90	4.30	5.60	6.10	14.00	6.60

[a]Values computed using two significant digits.
[b]French horn, bass/tenor trombones, trumpet, and tuba.
[c]Double bass, cello, viola, and violin.
[d]Flute, bass/alto flutes, bass/Bb/Eb clarinets, and alto/soprano saxophones.
[e]Double bass and violin.

were empirically chosen in a pilot test by decreasing their magnitude until no significant improvement in performance was observed. In all cases, the number of steps was forced to be at least as large as the number of steps used on the ERB scale.

Overall, no alternative feature made SWIPE′ improve neither consistently nor on average over all the databases, although some of the alternative features performed as well or almost as well as the proposed features, specifically, the parabolic and Gaussian kernels, and all window types. However, there were some features that made SWIPE′ improve significantly (by more than 10% relative to the GER) on the musical instruments database, specifically, the flat kernel envelope, and the Hertz, Barks, and mel scales. Inspection of the spectrum of the signals on which SWIPE′ improved showed that the spectrum was far from having the expected $1/f$ envelope. Instead, the spectrum tended to increase with harmonic number at low order harmonics, and then decreased after a relatively high order harmonic. This is consistent with the improvement produced by the flat envelope, which gives relatively more weight to high order harmonics than the $1/f$ envelope does. It is also consistent with the improvement produced by the Hertz and mel scales, which give a relatively better sampling of the high frequencies compared to the ERB scale [i.e., the transition from linear to logarithmic behavior in Eq. (5) occurs at larger values of σ]. Such statement is hard to make for the Bark scale since Eqs. (5) and (6) are not directly comparable. However, the GERs of Table VII suggest that the Bark scale behaves as being between the mel and the Hertz scales, probably because its frequency scaling factor (1960) is between the frequency scaling factors of the mel and Hertz scales (700 and infinity, respectively). Finally, the last row of Table VII shows the results of combining the flat envelope and the Hertz scale.[6] The combination of these features made SWIPE′ improve even more on the musical instruments database, but also worsen on the speech databases.

The discussion in the previous paragraph suggests that

TABLE VI. GERs for musical instruments by octave (MIS database).[a]

	Gross error (%)						
Algorithm	46.2 Hz +/−1/2 oct.	92.5 Hz +/−1/2 oct.	185 Hz +/−1/2 oct.	370 Hz +/−1/2 oct.	740 Hz +/−1/2 oct.	1480 Hz +/−1/2 oct.	Average
SWIPE′	1.20	1.00	2.30	0.89	0.13	0.29	0.97
SWIPE	0.08	1.20	3.00	1.00	0.25	0.38	0.99
YIN	3.20	0.95	5.30	1.80	0.69	0.96	2.20
AC-P	0.24	2.00	7.80	2.50	0.71	0.30	2.30
SHS	7.80	2.60	3.20	1.20	0.23	0.14	2.50
CC	0.26	2.60	8.20	2.70	0.93	0.40	2.50
TEMPO	15.00	2.80	2.00	1.10	0.52	0.31	3.60
ESRPD	7.90	2.60	4.80	4.20	12.00	32.00	11.00
SHR	37.00	0.60	1.80	27.00	70.00	81.00	36.00
Average	8.10	1.80	4.30	4.70	9.50	13.00	6.90

[a]Values computed using two significant digits.

TABLE VII. GERs of variations of SWIPE′ on PBD, KPD, DVD, and MIS databases.[a]

Variation	Gross error (%)				
	PBD	KPD	DVD	MIS	Average
Original	0.13	0.83	0.63	1.10	0.67
Flat envelope	0.16	1.00	1.40	0.60	0.79
Raw spectrum and $1/f$ envelope[b]	0.30	2.20	1.70	10.00	3.60
Squared spectrum and $1/f^2$ envelope[b]	12.00	14.00	9.70	22.00	14.00
Pulsed kernel	0.21	0.84	3.00	2.60	1.70
Parabolic kernel	0.13	0.83	0.62	1.10	0.67
Gaussian kernel	0.13	0.83	0.63	1.10	0.67
Hertz scale[c]	0.23	1.70	1.40	0.37	0.93
Bark scale[c]	0.18	1.00	0.90	0.61	0.67
Mel scale[c]	0.16	0.97	0.88	0.67	0.67
Logarithmic scale[c]	0.14	1.10	0.89	2.30	1.10
Fixed window size[d]	0.15	0.77	1.70	9.10	2.90
Hamming window ($k=2$)[e]	0.13	0.80	0.64	1.20	0.69
Blackman window ($k=3$)[e]	0.14	0.80	0.74	1.20	0.72
Gaussian window ($k=3.14$)[e]	0.13	0.83	0.63	1.10	0.67
Blackman-Harris window ($k=4$)[e]	0.15	0.76	0.77	1.20	0.72
Flat envelope and Hertz scale[f]	0.49	5.00	2.30	0.17	2.00

[a]Values computed using two significant digits.
[b]The use of the raw or squared spectrum implies the use of a kernel whose envelope decays as $1/f$ or $1/f^2$, respectively, to match the spectral envelope of a sawtooth waveform.
[c]Spectrum was computed using FFTs and was inter/extrapolated to equidistant steps in the specified scale (see Table VIII for step size).
[d]The power-of-2 window size whose optimal pitch was closest to the geometric mean pitch of the database was used in each case (1024 for the speech databases and 256 for the musical instruments database).
[e]Window type was selected among all the window types with that value of k for having the spectrum whose square root produced the largest normalized inner product between its main lobe and one period of the cosine around zero (see Table I).
[f]The spectral analysis was limited to the range 0–5 kHz (see text for details).

SWIPE and SWIPE′ would benefit from preprocessing the signal with an auditory model, as in the work by Klapuri (2008). Klapuri (2008) showed that preprocessing the signal with a gammatonelike filterbank, followed by a half-wave rectifier, a compressive function, and a low pass filter, tend to produce low order harmonics in each of the outputs of the system, even in the outputs that came from filters that responded only to high order harmonics. This means that such preprocessing would boost (or make appear) weak (or missing) low order harmonics, making the signal more suitable for analysis with SWIPE or SWIPE′. However, this claim needs further validation.

IV. SUMMARY AND CONCLUSIONS

SWIPE estimates the pitch as the fundamental frequency of the sawtooth waveform whose spectrum best matches the spectrum of the input signal. The schematic description of the algorithm is the following.

(1) For each pitch candidate f within a pitch range $f_{\min}-f_{\max}$, compute its score as follows:
   (a) Compute the square root of the spectrum of the signal.
   (b) Normalize the square root of the spectrum and apply an integral transform using a normalized cosine kernel whose envelope decays as $1/\sqrt{f}$.

TABLE VIII. Step sizes and number of steps used to sample the spectrum for each of the frequency scales and databases. The range used to sample the spectrum was lower bounded by one-quarter of the lowest expected pitch and upper bounded by the frequency specified in parenthesis (the Nyquist frequency on each case). (DVD has two columns because it contains files sampled at two different sampling rates.)

Scale	Step size	Number of steps			
		MIS (5 kHz)	PBD & KPD (10 kHz)	DVD (12.5 kHz)	DVD (25 kHz)
ERBs	1/10	288	350	370	434
Bark	1/20	384	446	461	495
Mel	5	471	612	659	809
Logarithmic	1/96 octave	901	957	988	1084
Hertz	5	1000	2000	2500	5000

(2) Estimate the pitch as the highest score candidate.

An implicit objective of the algorithm is to find the frequency for which the average peak-to-valley distance at its harmonics is maximized. To achieve this, the kernel is set to zero below the first negative lobe and above the last negative lobe, and to avoid bias, the magnitude of these two lobes is halved. To make the contribution of each harmonic of the sawtooth waveform proportional to its amplitude, the square root of the spectrum is taken before applying the integral transform. To make the kernel match the normalized square-root spectrum of the sawtooth waveform, a $1/\sqrt{f}$ envelope is applied to the kernel, which is normalized using only its positive part, and the spectrum is computed using a window size that makes the width of the main spectral lobes match the width of the lobes of the kernel.

Several techniques are applied to reduce computational cost. First, the optimal window size is replaced with the two closest power-of-2 window sizes, for which it is more efficient to compute a FFT, and the scores are appropriately combined to produce a single score. This approach has the extra advantage of allowing a FFT to be shared by several pitch candidates. Second, the scores are computed using a coarse resolution and then fine-tuned using parabolic interpolation. Third, the window overlap is minimized while allowing the pitch of a signal as short as four cycles to be recognized. Last, the inner product between the kernel and the square root of the spectrum is computed on the ERB frequency scale, since this scale emphasizes the regions where most of the spectral energy is concentrated. This last technique not only reduces the computational cost but also improves the performance (as shown in the tests).

SWIPE′, a variation of SWIPE, uses only the first and prime harmonics of the signal, which largely decreases the scores of subharmonics of the pitch, significantly reducing the chances of estimating the pitch as one of its subharmonics.

SWIPE and SWIPE′ were tested using speech and musical instruments databases and their performance was compared against 12 other algorithms, which have been cited in literature and for which free implementations exist. SWIPE′ was shown to outperform all the algorithms on all the databases. SWIPE was ranked second in the normal speech and musical instruments databases, and was ranked third in the disordered speech database.

APPENDIX: EXTRA DETAILS OF THE EVALUATION

1. Disordered voice database

The fundamental frequency estimates included in the disordered voice database were not used as ground truth. The reason was that we wanted to estimate pitch, and fundamental frequency does not always correspond to pitch, as mentioned in the Introduction.

Since this database consisted of sustained vowels and most of them had a relatively stable pitch, we used a subject with vast experience in music transcription (the first author) to label the samples with their pitch, by matching them to the closest note playing sawtooth waveforms on an electronic keyboard. Assuming that he correctly chose one of the two closest notes, this procedure should introduce an error no larger than 6% (one semitone), which is smaller than the 20% necessary to produce a gross error (see its definition in Sec. III C).

There are some samples in this database for which the pitch spans a perfect fourth or more (i.e., the highest pitch is more than 33% higher than the lowest pitch). Since this range is relatively large compare to the permissible error ($\pm 20\%$), these samples were excluded. Only samples for which the range did not span more than a major third (i.e., the highest pitch does not exceed the lowest pitch by more than 26%) were preserved, and they were assigned the pitch of the note corresponding to the median of the range. If the median was between two notes, the pitch of any of them was assigned to the sample. This should introduce an error no larger than two semitones (12%), which is about half the maximum permissible error of 20%. There were 30 samples for which the subject could not confidently perceive pitch; hence, those samples were also excluded.

The GERs on this database were first computed per sample (vowel), and then averaged over the samples. Since the ground truth data were based on the perception of only one listener, it could be argued that these data have low validity. To alleviate this, we excluded the samples for which the minimum GER was larger than 50%.

2. Musical instruments database

The files were downsampled from 44.1 to 10 kHz in order to reduce computational cost. No noticeable change of pitch was perceived by doing this, even for the highest pitch sounds.

The files were labeled assuming that their content was a chromatic scale. However, some of the intervals were imprecise in some of the files, leading to accumulated errors that exceeded one semitone, and consequently to wrong labels. This situation was especially common among string instruments, especially when playing *pizzicato*. To correct this situation, the first author listened to the files and relabeled them using an electronic keyboard as reference. This procedure introduced repeated file names, which were removed by keeping only the sounds whose pitch was closest to the target. When the conflicting files had notes whose pitches were equally close to the target, the file with the best sound quality was preserved. This removal of files was done to avoid the overhead of having to add extra symbols to the file names to allow for repetitions, which would have complicated the generation of scripts to test the algorithms. Since the process of manually correcting the names of the notes was very tedious, especially for pizzicato sounds, only the labels of bass and violin pizzicato files were fixed, and the cello and viola pizzicato sounds were excluded from the evaluation.

The commands issued for each of the algorithms were the following.

AC-P. To pitch (ac)…0.001 30 15 no 0.03 0.45 0.01 0.35 0.14 1666.

CC. To Pitch (cc)…0.001 30 15 no 0.03 0.45 0.01 0.35 0.14 1666.

ESRPD. pda input_file -o output_file -P -d 1 -shift 0.001 -length 0.0384 -fmax 1666 -fmin 30 -n 0 -m 0.

SHS. To pitch (shs)… 0.001 30 15 5000 15 0.84 1666 48.

SHR. [t,p]=shrp(x,fs,[30 1666],40,1,0.4,5000,0,0).

SWIPE′. [p,t]=swipep(x,fs, [30 1666], 0.001, 1/96, 0.1,-Inf).

YIN. p.minf0=30; p.maxf0=1666; p.hop=10; p.sr=10000; r=yin(x,p).

Since the range 30–1666 Hz was too large for the Speech Filing System algorithms (AC-S, ANAL, CEP, and RAPT), they were not evaluated on this database.

The GERs on this database were first computed per sample (i.e., note) and then averaged over the samples. However, there were some samples for which agreement on existence of pitch among the algorithms existed only at very few instants of time (only one in some cases). To avoid giving too much emphasis to these few instants, only samples in which the algorithms agreed that pitch existed in more than 50% of the time were used in the statistics.

3. Speech databases (KPD and PBD)

The commands issued for each of the algorithms were the following[7]

AC-P. To pitch (ac)…0.001 40 15 no 0.03 0.45 0.01 0.35 0.14 800.

AC-S. fxac input_file.

ANAL. fxanal input_file.

CC. To pitch (cc)…0.001 40 15 no 0.03 0.45 0.01 0.35 0.14 800.

CEP. fxcep input_file.

ESRPD. pda input_file -o output_file -L -d 1 -shift 0.001 -length 0.0384 -fmax 800 -fmin 40 -lpfilter 600.

RAPT. fxrapt input_file.

SHS. To pitch (shs)…0.001 40 15 1250 15 0.84 800 48.

SHR. [t,p]=shrp(x,fs,[40 800],40,1,0.4,1250,0,0).

SWIPE. [p,t]=swipe(x,fs,[40 800], 0.001, 1/96, 0.1,-Inf).

SWIPE′. [p,t]=swipep(x,fs,[40 800], 0.001, 1/96, 0.1, -Inf).

TEMPO. f0raw=exstraightsource(x,fs).

YIN. p.minf0=40; p.maxf0=800; p.hop=20; p.sr=fs; r=yin(x,p).

[1]This assertion can be inferred from the analysis by Camacho and Harris (2007).

[2]In this work, we will make use of the magnitude of the spectrum but not its phase. Hence, to abbreviate, the words *magnitude of* will be omitted, and whenever the word *spectrum* is found, it should be understood as *magnitude of the spectrum*.

[3]A normalized inner product equal to 1 is unattainable since side lobes will inevitably appear in the same region as the negative part of the cosine.

[4]It was downsampled to 10 kHz in our evaluation to save computational cost (see Appendix for details).

[5]The logarithm of the spectrum was not evaluated because it can be negative, which would make the normalization in Eq. (12) nonsense. Also, only the window types with the largest normalized inner product between the kernel and the square root of the spectrum evaluated over the whole period of the cosine for each value of k in Table I were evaluated. The flattop window was left out because its high value of k made its computational cost too high for our evaluation.

[6]For this test, the spectral analysis was limited to 5 kHz in *all* databases. This was done to discard the possibility that the use of a higher upper limit of frequency for speech (10–25 kHz) than for musical instruments (5 kHz) may have had an effect on the performance divergence (i.e., increase of performance on musical instruments but decrease of performance on speech) occurred in the tests where the flat envelope and the Hertz scales were evaluated separately.

[7]The command for CATE is not reported because we used our own implementation of the algorithm.

American Standards Association (**1960**). "Acoustical Terminology SI 1-1960," American Standards Association, New York.

Bagshaw, P. C., Hiller, S. M., and Jack, M. A. (**1993**). "Enhanced pitch tracking and the processing of F0 contours for computer aided intonation teaching," *Proceedings of the Third European Conference on Speech Communications and Technology*, pp. 1003–1006.

Bagshaw, P. C. (**1994**). "Automatic prosodic analysis for computer aided pronunciation teaching," Ph.D. thesis, University of Edinburgh, Edinburgh.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proceedings of the Institute of Phonetic Sciences **17**, 97–110.

Camacho, A., and Harris, J. G. (**2007**). "A pitch estimation algorithm based on the smooth harmonic average peak-to-valley envelope," Proceedings of the International Symposium on Circuits and Systems, pp. 3940–3943.

Dannenberg, R. B., Birmingham, W. P., Tzanetakis, G. P., Meek, C. P., Hu, N. P., and Pardo, B. P. (**2004**). "The MUSART testbed for query-by-humming evaluation," Comput. Music J. **28**, 34–48.

De Bot, K. (**1983**). "Visual feedback of intonation I: Effectiveness and induced practice behavior," Lang Speech **26**, 331–350.

de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.

Di Martino, J., and Laprie, Y. (**1999**). "An efficient *F*0 determination algorithm based on the implicit calculation of the autocorrelation of the temporal excitation signal," Proceedings of the Sixth European Conference on Speech Communication and Technology, pp. 2773–2776.

Doughty, J., and Garner, W. (**1947**). "Pitch characteristics of short tones. I. Two kinds of pitch threshold," J. Exp. Psychol. **37**, 351–365.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am. **71**, 1568–1580.

Fant, G. (**1970**). *Acoustic Theory of Speech Production, With Calculations Based on X-Ray Studies of Russian Articulations* (Mouton De Gruyter, The Hague).

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

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).

Kawahara, H., Katayose, H., de Cheveigné, A., and Patterson, R. D. (**1999**). "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity," Proceedings of the Sixth European Conference on Speech Communication and Technology, pp. 2781–2784.

Klapuri, A. (**2004**). "Automatic music transcription as we know it today," J. New Music Res. **33**, 269–282.

Klapuri, A. (**2008**). "Multipitch analysis of polyphonic music and speech signals using an auditory model," IEEE Trans. Audio, Speech, Lang. Process. **16**, 255–266.

Medan, Y., Yair, E., and Chazan, D. (**1991**). "Super resolution pitch determination of speech signals," IEEE Trans. Signal Process. **39**, 40–48.

Meddis, R., and Hewitt, M. J. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Noll, A. M. (**1967**). "Cepstrum pitch determination," J. Acoust. Soc. Am. **41**, 293–309.

Oppenheim, A. V., Schafer, R. W., and Buck, J. R. (**1999**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Patel, A. D., and Balaban, E. (**2001**). "Human pitch perception is reflected in the timing of stimulus-related cortical activity," Nat. Neurosci. **4**, 839–844.

Plante, F., Meyer, G., and Ainsworth, W. A. (**1995**). "A pitch extraction

reference database," Proceedings of the Fourth European Conference on Speech Communication and Technology, pp. 837–840.
Rabiner, L. R. (**1977**). "On the Use of Autocorrelation Analysis for Pitch Detection," IEEE Trans. Acoust., Speech, Signal Process. **25**, 24–33.
Schroeder, M. R. (**1968**). "Period histogram and product spectrum: New methods for fundamental frequency measurement," J. Acoust. Soc. Am. **43**, 829–834.
Schwartz, D. A., and Purves, D. (**2004**). "Pitch is determined by naturally occurring periodic sources," Hear. Res. **194**, 31–46.
Secrest, B., and Doddington, G. (**1983**). "An integrated pitch tracking algorithm for speech systems," Proceedings of ICASSP-83, pp. 1352–1355.
Sondhi, M. M. (**1968**). "New methods of pitch extraction," IEEE Trans. Audio Electroacoust. **AU-16**, 262–266.
Spanias, A. S. (**1994**). "Speech coding: A tutorial review," Proc. IEEE **82**, 1541–1582.
Sun, X. (**2000**). "A pitch determination algorithm based on subharmonic-to-harmonic ratio," Proceedings of the International Conference on Spoken Language Processing, Vol. **4**, pp. 676–679.
Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.
Wang, M., and Lin, M. (**2004**). "An analysis of pitch in Chinese spontaneous speech," International Symposium on Tonal Aspects of Tone Languages, Beijing, China.
Yumoto, E., Gould, W. J., and Baer, T. (**1982**). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. **71**, 1544–1549.

Harmonic segregation through mistuning can improve fundamental frequency discrimination

Joshua G. W. Bernstein[a)]
Army Audiology and Speech Center, Walter Reed Army Medical Center, 6900 Georgia Avenue N.W., Washington, DC 20307

Andrew J. Oxenham
Department of Psychology, 75 East River Parkway, University of Minnesota, Minneapolis, Minnesota 55455

(Received 14 October 2007; revised 16 June 2008; accepted 17 June 2008)

This study investigated the relationship between harmonic frequency resolution and fundamental frequency (f_0) discrimination. Consistent with earlier studies, f_0 discrimination of a diotic bandpass-filtered harmonic complex deteriorated sharply as the f_0 decreased to the point where only harmonics above the tenth were presented. However, when the odd harmonics were mistuned by 3%, performance improved dramatically, such that performance nearly equaled that found with only even harmonics present. Mistuning also improved performance when alternating harmonics were presented to opposite ears (dichotic condition). In a task involving frequency discrimination of individual harmonics within the complexes, mistuning the odd harmonics yielded no significant improvement in the resolution of individual harmonics. Pitch matches to the mistuned complexes suggested that the even harmonics dominated the pitch for f_0's at which a benefit of mistuning was observed. The results suggest that f_0 discrimination performance can benefit from perceptual segregation based on inharmonicity, and that poor performance when only high-numbered harmonics are present is not due to limited peripheral harmonic resolvability. Taken together with earlier results, the findings suggest that f_0 discrimination may depend on auditory filter bandwidths, but that spectral resolution of individual harmonics is neither necessary nor sufficient for accurate f_0 discrimination. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956484]

PACS number(s): 43.66.Hg, 43.66.Mk, 43.66.Fe, 43.66.Ba [RLF] Pages: 1653–1667

I. INTRODUCTION

The ability of human listeners to discriminate small differences in pitch, as estimated by the f_0 difference limen (f_0 DL), is typically best when at least some low-numbered harmonics are present (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994). Studies that have measured f_0 DLs as a function of the lowest harmonic present have usually found an abrupt transition between good and poor performance as the lowest harmonic number is increased from around 9 to 12 (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003), at least for f_0's in the adult speech range, between about 100 and 200 Hz. The reasons underlying the dependence of f_0 discrimination on harmonic number are not fully understood, although most emphasis has been placed on differences in peripheral resolvability between low- and high-numbered harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham 2006a, b).

There is some evidence suggesting a link between peripherally resolved harmonics and good pitch perception. First, the relatively abrupt increase in f_0 DL with increasing lowest harmonic number typically matches the point at which effects of component phases are first observed (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2006a, b). Phase effects indicate interactions between neighboring components, suggesting that the harmonics are at least partially unresolved at the point where f_0 DLs increase. Second, recent studies have demonstrated a relationship between the minimum harmonic spacing required for relatively accurate f_0 discrimination performance and auditory filter bandwidth in hearing-impaired listeners (Bernstein and Oxenham, 2006b), as well as in normal-hearing listeners at different sound intensities (Bernstein and Oxenham, 2006a). Because auditory filter bandwidth, relative to harmonic spacing, is thought to determine the degree to which harmonics are resolved, the link between filter bandwidth and f_0 DLs could be viewed as evidence that resolved harmonics are necessary for good pitch perception.

On the other hand, there are some grounds to question the relationship between harmonic resolvability and f_0 DLs. For instance, it has been found that increasing the number of peripherally resolved harmonics by presenting successive harmonics to alternating ears (i.e., the odd harmonics to one ear and the even harmonics to the other) does not improve f_0 DLs or pitch identification performance for stimuli consisting of either two (Houtsma and Goldstein, 1972; Arehart and Burns, 1999) or many (Bernstein and Oxenham, 2003) harmonics. Thus, a reasonable summary of the results so far might be that resolved harmonics seem to be *necessary* (Houtsma and Smurzynski, 1990; Shackleton and Carlyon,

a)Author to whom correspondence should be addressed. Electronic mail: joshua.bernstein@amedd.army.mil

1994; Bernstein and Oxenham 2006a, b) but not *sufficient* (Houtsma and Goldstein, 1972; Bernstein and Oxenham, 2003) for good pitch perception.

An alternative interpretation is that f_0 DLs are not related to harmonic resolvability at all, but only to auditory filter bandwidth. De Cheveigné and Pressnitzer (2006) illustrated this point in their recent model, which assumes that the coding of f_0 within a given auditory filter depends on the relationship between the stimulus period $(1/f_0)$ and the duration of the impulse response of the auditory filter. Specifically, if the impulse response of the filter is shorter than the period of the waveform, the f_0 will not be well represented. Because the impulse response is inversely related to filter bandwidth, which in the case of the auditory system is roughly proportional to center frequency, it follows that the accuracy of f_0 coding will decrease with increasing harmonic number (de Cheveigné and Pressnitzer, 2006). This type of approach can in principle account for why f_0 DLs are affected by changes in filter bandwidth, without being dependent on harmonic resolvability *per se*. In other words, according to de Cheveigné and Pressnitzer's (2006) formulation, peripherally resolved harmonics are neither necessary nor sufficient for good pitch perception.

The aim of the present study was to provide an empirical test of the link between harmonic resolvability and pitch perception, with the aim of dissociating peripheral resolvability from harmonic number and filter bandwidth. Our starting point was the finding by Bernstein and Oxenham (2003) that presenting successive harmonics to alternating ears improved the peripheral resolvability of individual harmonics, but did not affect f_0 DLs. This result is puzzling because, in principle, listeners could have attended just to the ear with the even harmonics, and extracted a pitch corresponding to twice the nominal f_0, with half the harmonic number (e.g., the 12th harmonic of 100 Hz could have been interpreted as the 6th harmonic of 200 Hz). The fact that listeners were not able to utilize this strategy could be interpreted in at least two ways. First, pitch may be extracted from a representation in which the input from each ear is automatically fused, forming an unresolved "central spectrum" (Zurek, 1979); in this case harmonics would need to be resolved in both the monaural and binaural sense. Second, the harmonics may be resolved within the central representation, but the odd harmonics may inhibit activation of a harmonic template at twice the f_0. According to this interpretation, the pitch of a complex sound is extracted by selecting the centrally stored harmonic template that best matches the set of individual component frequencies present in the input stimulus (Goldstein, 1973). Presumably both a 100- and 200-Hz template would be activated by a diotic 100-Hz tone complex. However, spurious octave confusions might be avoided if the activation of the 200-Hz template were reduced by inhibitory inputs at intermediate component frequencies associated with the odd harmonics of the 100-Hz tone complex. In the case of a dichotic 100-Hz complex, the odd components in one ear might still inhibit the 200-Hz template that would otherwise be activated by the even components in the other ear.

It is likely that this putative inhibition mechanism is susceptible to auditory grouping and segregation constraints, in order to avoid interference between simultaneous complex tones with different f_0's. Just as the perceptual removal of a frequency component via mistuning or onset disparity reduces that component's contribution to the pitch of a tone complex (e.g., Moore *et al.*, 1985; Darwin and Ciocca, 1992; Darwin *et al.*, 1995), such manipulations might be predicted to reduce its role in inhibiting the activation of a particular harmonic template. Auditory grouping by harmonicity may explain the contrast between the results of Bernstein and Oxenham (2003) and those of Beerends and Houtsma (1986), who showed that listeners were able to independently identify the f_0 of two two-tone complexes presented to opposite ears. The fact that the harmonics presented to opposite ears in the Bernstein and Oxenham (2003) study shared the same fundamental frequency may have encouraged their perceptual fusion.

In this study, we introduced constant but small mistunings of all the odd harmonics to encourage their perceptual segregation from the even harmonics, as has been done for individual harmonics in the past (e.g., Moore *et al.*, 1985; Darwin and Ciocca, 1992; Darwin *et al.*, 1995). We tested both diotic (all harmonics to both ears) and dichotic (odd and even harmonics to opposite ears) conditions. The predictions were as follows: (1) if f_0 DLs depend on harmonic resolvability within a "central spectrum" (Zurek, 1979), then a slight mistuning of the odd harmonics should have little or no effect on resolvability, and so should not affect f_0 DLs at all; (2) if f_0 DLs depend on monaural (peripheral) harmonic resolvability, but mistuning is sufficient to perceptually segregate components presented to opposite ears, then the mistuning should improve f_0 discrimination in the dichotic condition but not in the diotic condition, where performance is limited by peripheral resolvability of the harmonics; (3) if f_0 DLs do not depend on harmonic resolvability at all, but rather on the harmonic number associated with the perceived f_0 (or filter bandwidths), then mistuning the odd harmonics should improve performance in both diotic and dichotic conditions by facilitating the perceptual segregation of the odd and even harmonics.

II. EXPERIMENT 1: FUNDAMENTAL FREQUENCY DISCRIMINATION

A. Methods

1. Stimuli

This experiment measured f_0 discrimination performance as a function of f_0 for diotic and dichotic harmonic complexes and complexes in which the f_0 of the odd harmonics was shifted relative to that of the even harmonics. All complexes were bandpass filtered into a fixed spectral region (Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2005; 2006a, b). This paradigm was employed instead of fixing the f_0 and varying the harmonic number (e.g., Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003) for two reasons. First, in a fixed-f_0 paradigm, an alternative cue to f_0 discrimination exists based on the spectral region occupied by the components presented. While the usefulness of this cue can be diminished by applying a random rove to the lowest harmonic presented from interval to interval

(Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003; Moore *et al.*, 2006), such a cue can still influence the f_0 DL estimate in the case of large f_0 DLs associated with high-order harmonics (Bernstein and Oxenham, 2003; Moore *et al.*, 2006). Second, the fixed bandpass-filter paradigm allows an estimate of the f_0 needed to achieve a fixed level of f_0 discrimination performance via an adaptive tracking technique (e.g., Krumbholz *et al.*, 2000). Because this technique was used in experiment 2 to determine the size of the mistuning required to induce a change in f_0 discrimination accuracy, experiment 1 also used a bandpass filter paradigm, to maintain similar stimulus parameters across the two experiments. Experiment 1A measured f_0 discrimination performance for diotic and dichotic, harmonic and f_0-shifted complexes. Experiment 1B provided a control that compared f_0 discrimination performance for complexes consisting of both even and odd harmonics with performance for complexes consisting of only odd harmonics.

The stimuli and methods in this experiment were similar to those employed by Bernstein and Oxenham (2006a, b). The stimuli consisted of 300-ms (including 30-ms raised cosine rise and fall ramps) bandpass-filtered random-phase tone complexes. A new set of phases was selected independently from a uniform distribution for each 300-ms tone complex. The bandpass filter was held constant throughout the experiment, with 1.5- and 3.5-kHz corner frequencies and 50 dB/octave low- and high-frequency slopes. The filtering operation was implemented in the spectral domain by first adjusting the amplitude of each sinusoidal component, then summing all the components together.

Tone complexes were constructed by selecting even and odd harmonics from the same f_0 (harmonic conditions) or from two different f_0's (f_0-shifted conditions), whereby the f_0 for the odd components ($f_{0,\mathrm{odd}}$) was 3% higher than the f_0 for the even component frequencies ($f_{0,\mathrm{even}}$). In the remainder of the paper the f_0 reported is the $f_{0,\mathrm{even}}$. Even and odd harmonics were presented either diotically (all components to both ears) or dichotically (even and odd components to opposite ears), for a total of four conditions. For all conditions, filtered complexes consisting of only even or only odd components were synthesized separately. In the diotic conditions, the even and odd complexes were summed together and presented to both ears simultaneously. In the dichotic conditions, odd and even components were randomly assigned to the right and left ears on a trial-by-trial basis. Fundamental frequency discrimination was tested for seven average values of f_0 (50, 75, 100, 125, 150, 175, and 200 Hz).

To reduce the effectiveness of loudness as an alternative discrimination cue, the root-mean-squared amplitude of the combined even components was first equalized across the three intervals and then a random level perturbation was added to each interval, chosen from a uniform distribution of ± 2.5 dB. The odd components were scaled to maintain the same level per component as the even components in each interval. In addition, $f_{0,\mathrm{base}}$ was roved from trial to trial within a run, chosen from a uniform distribution between $\pm 5\%$ of the average f_0. This was intended to encourage listeners to compare the pitches of the stimuli in each of the intervals of one trial, rather than comparing the pitch of each interval with some internally stored representation of the $f_{0,\mathrm{base}}$, although the f_0 roving may not have been effective for low f_0's where measured f_0 DLs were relatively large.

All stimuli were presented in a threshold equalizing noise (TEN; Moore *et al.*, 2000), lowpass filtered at 10 kHz and set at a level of 40 dB sound pressure level (SPL) per equivalent rectangular bandwidth ($\mathrm{ERB_N}$; Glasberg and Moore, 1990), which reduced the possibility of the use of combination tones and off-frequency listening. The noise was presented diotically to reduce the possibility of a binaural signal-detection advantage in the diotic stimulus conditions. The 0-dB sensation level (SL) reference for each individual listener was defined as the average detection threshold for 1.5-, 2.5- and 3.5-kHz probe tones presented monaurally to the left ear in monaural TEN (range: 35.4–37.1 dB SPL across the seven listeners who participated in experiments 1 through 4). Each equal-amplitude component (before filtering, where applicable) was presented at an average 12.5 dB SL (adjusted for each listener). With the 50 dB/octave filter slopes, only tones falling within about $\frac{1}{4}$ octave above and below the upper and lower filter cutoff frequencies, respectively, (1261–4162 Hz) were audible.

2. Procedure

Fundamental frequency DLs were estimated in a three-interval three-alternative forced-choice (3I-3AFC) adaptive procedure, using a two-down, one-up algorithm to track the 70.7% correct point on the psychometric function (Levitt, 1971). The three intervals were separated by gaps of 300 ms. The background noise (TEN) was gated on 200 ms before the first interval and gated off 100 ms after the third interval, producing a total noise duration of 1800 ms in each trial.

Two intervals contained a stimulus with a base f_0 ($f_{0,\mathrm{base}}$) and the other interval contained a complex with a higher f_0. The listener's task was to identify the interval containing the complex with the higher pitch. Feedback (correct/incorrect) was provided following each response. The f_0 difference (Δf_0), which was initially set to 15.8% of the f_0, changed by factors of 1.59 and 1.26 until the second and fourth reversals, respectively, and then changed by a factor of 1.12 for six more reversals. The f_0 DL was estimated as the geometric mean of the Δf_0's at the last six reversal points. Measurements were repeated four times for each f_0 and condition, for a total of 112 runs per listener. Each listener received at least 2 h of practice before data collection began.

3. Listeners

A list of the seven listeners that participated in the study (four of whom participated in experiment 1A) is shown in Table I, which includes information about age, gender, musical background, and the experiments in which each listener participated. All listeners had normal hearing (15 dB hearing level or less *re* ANSI-1996 at octave frequencies between 0.25 and 8 kHz). The five musicians each had at least four years of formal training.

TABLE I. Details of the normal-hearing listeners who participated in the study.

				Experiment					
Listener	Age	Gender	Musician	1A	1B	2	3	4A	4B
1	20	M	No	X		X	X		
2	31	F	Yes	X	X	X	X	X	X
3	28	F	No	X		X	X	X	
4	51	F	Yes	X		X	X	X	
5	36	F	Yes		X				X
6	22	M	Yes		X			X	X
7	24	M	Yes		X				X

4. Control experiment

A control experiment (1B) provided a baseline measure for the performance that would be expected if f_0 discrimination was based on the even or odd harmonics alone. Fundamental frequency DLs were measured in four listeners (Table I) as a function of f_0 for diotic harmonic complexes containing all harmonics or odd harmonics only, bandpass filtered as described above. Performance for complexes containing only even harmonics were inferred from the all-harmonics conditions with $2f_0$ (e.g., the 200-Hz all-harmonics condition was equivalent to a 100-Hz even-harmonics-only condition). Measurements were performed for the same seven f_0's (50–200 Hz) as experiment 1A, plus three additional f_0's (37.5, 250, and 300 Hz) that allowed additional comparisons between the all- and odd-harmonics conditions and the inferred even-harmonics situation. Each listener received at least 2 h of practice before data collection began.

B. Results

The mean results from experiment 1A are shown in Fig. 1. Geometric-mean f_0 DLs across the four listeners are plotted as a function of f_0 for the four experimental conditions, with error bars indicating the standard error. Results for each individual listener were generally consistent with the mean results and are not shown. These results are discussed along with the results of a repeated-measures analysis of variance (RMANOVA) with three within-listener factors: f_0, f_0 shift of the odd harmonics (0 or 3%) and mode of presentation (diotic or dichotic). The reported degrees of freedom throughout this study reflect a Huynh-Feldt (1976) correction that was applied wherever necessary.

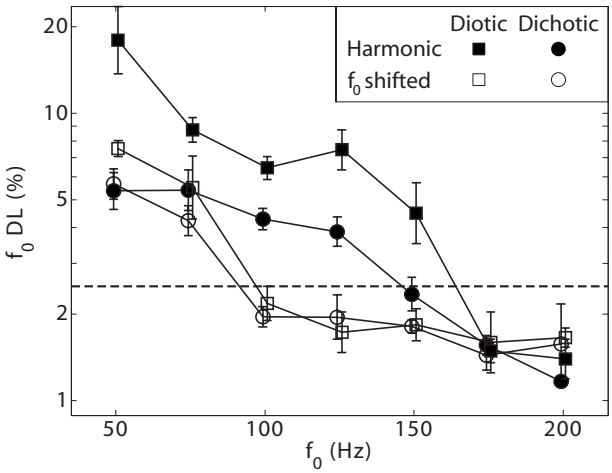


FIG. 1. Mean results of experiment 1A across four listeners. Estimated f_0 DLs are plotted function of f_0 for bandpass filtered tone complexes. Error bars indicate ± one standard error across the four listeners. The dashed line indicates an f_0 DL of 2.5%, the fixed f_0 difference used in experiment 2. An f_0 discrimination benefit was obtained by shifting the f_0 of the odd harmonics for both diotic and dichotic stimuli. A smaller f_0 discrimination benefit was also obtained for harmonic stimuli under dichotic presentation.

Three key findings are apparent in the data. First, f_0 DLs tended to decrease (improve) with increasing f_0, consistent with previous findings (e.g., Bernstein and Oxenham, 2005), as confirmed by a significant main effect of f_0 [$F(6,18)=25.7$, $p<0.0005$]. Second, the f_0 at which the transition from relatively large (4% or more) to relatively small f_0 DLs (less than 2.5%, horizontal dashed line) occurred was approximately an octave higher in the harmonic conditions than in the f_0-shifted conditions, consistent with a significant main effect of f_0 shift [$F(1,3)=83.8$, $p<0.005$]. Third, in the harmonic conditions, dichotic presentation yielded smaller f_0 DLs for those f_0's that yielded relatively poor performance (f_0 DLs > 4%) under diotic presentation, as suggested by a significant main effect of diotic versus dichotic presentation [$F(1,3)=36.2$, $p<0.01$]. Because the interaction between f_0 shift and mode of presentation was not significant [$F(1,3)=1.27$, $p=0.34$], an effect of dichotic presentation in the f_0-shifted conditions cannot be ruled out, although it was only visually apparent in harmonic conditions. A significant three-way interaction between f_0, f_0 shift, and mode of presentation [$F(2.1,6.3)=5.7$, $p<0.05$] is consistent with the observation that the benefit of dichotic presentation was mainly observed in harmonic, low-f_0 conditions, although there was also a small effect at the lowest f_0's for the f_0-shifted conditions.

The results of control experiment 1B are shown in Fig. 2. Geometric-mean f_0 DLs across the four listeners are plotted as a function of f_0 for the all-harmonics (squares) and odd-harmonics only conditions (diamonds). The all-harmonics data are replotted (triangles) at half the f_0 to represent the even-harmonics only condition. The odd-harmonics data were compared to the all-harmonics and even-harmonics conditions in two separate RMANOVAs, each with two factors: f_0 and condition (odd versus all or odd vs. even harmonics). While all ten f_0's were included in the odd- versus all-harmonic RMANOVA, only the six f_0's that

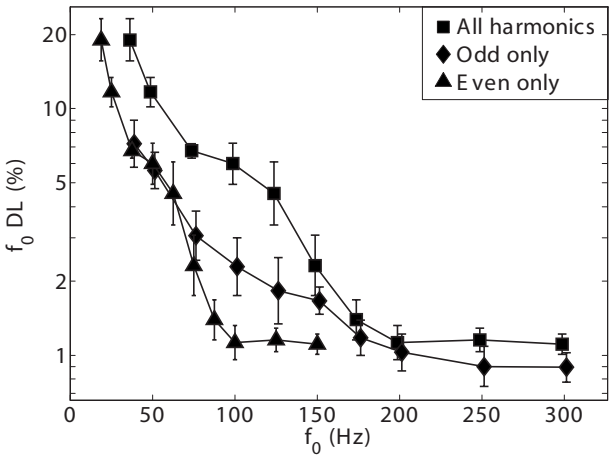


FIG. 2. Mean results of experiment 1B across four listeners. Estimated f_0 DLs are plotted as a function of f_0 for bandpass filtered tone complexes containing all harmonics or odd harmonics only. The all-harmonics data are replotted as an even-harmonics only condition by dividing the f_0 by two. Error bars indicate ± one standard error across the four listeners.

were represented in both conditions were included in the odd-versus even-harmonics comparison (37.5, 50, 75, 100, 125, and 150 Hz).

Both analyses showed a main effect of f_0 [odd versus all: $F(4.7,14.1)=83.8$, <0.0005; odd versus even: $F(5,15)=34.8$, $p<0.0005$], reflecting the improvement in f_0 DLs observed with increasing f_0. In contrast to previous studies, no clear plateau in performance is reached at very low f_0's. This is probably due to our use of very low f_0's, extending down to values that are close to the lower absolute limits of pitch perception (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001).

Presenting only the odd harmonics yielded a clear benefit to f_0 discrimination performance: There was a significant main effect of condition in the odd versus all comparison [$F(1,3)=43.6$, $p<0.01$]. Nevertheless, the improvement yielded by presenting the odd harmonics alone was not as great as the improvement yielded by presenting the even harmonics alone: There was also a significant main effect of condition in the odd versus even comparison [$F(1,3)=20.6$, $p<0.05$]. There was a significant interaction between f_0 and condition in the all versus odd comparison [$F(9,27)=5.1$, $p<0.0005$], reflecting the observation that the improvement yielded by presenting the odd harmonics only occurred for low and not high f_0's. Although the difference between the odd and even conditions was visually apparent only for higher f_0's, the interaction between f_0 and condition did not reach significance [$F(4.0,12.1)=2.78$, $p=0.08$]. This may be due in part to the reduced number of f_0's available for this comparison and the relatively low statistical power afforded by four subjects.

C. Discussion

The main finding of this experiment is that shifting the f_0 of the odd harmonics with respect to the even harmonics improved f_0 discrimination at low f_0's. For the diotic condition (Fig. 1; squares), the results with the shifted components (open squares) closely matched those with the purely harmonic components (filled squares), when the f_0's of the former were doubled. In other words, performance in the shifted condition was as good as if the odd harmonics were completely absent. In terms of the three hypotheses laid out in the introduction, this basic result seems to rule out the first, "central spectrum" hypothesis (Zurek, 1979). This is because the 3% shift in the frequencies of the odd harmonics is unlikely to increase the peripheral resolvability of the harmonics: Although the shift increased the frequency spacing between each even and the higher adjacent odd harmonic, it also decreased the spacing between the even harmonic and its lower adjacent odd harmonic, resulting in no predicted gain in resolvability. This prediction is tested empirically in experiment 3.

The second finding is that the improvement in f_0 discrimination produced by f_0-shifted odd harmonics was found for both the diotic and dichotic conditions. This finding seems to rule out the second hypothesis presented in the introduction, that peripherally resolved harmonics are necessary for good pitch discrimination: As discussed above (and addressed in experiment 3) shifting the f_0 of the odd harmonics is unlikely to have increased the peripheral resolvability of the harmonics, but still led to an improvement in f_0 discrimination.

The results of experiment 1B (Fig. 2) provide further evidence in support of the idea that f_0 discrimination performance is not governed by peripheral resolvability. Improving resolvability by removing the even harmonics did not yield the small ($<2\%$) f_0 DLs generally associated with low order harmonics that was observed for the f_0-shifted conditions of experiment 1A, or for the even-harmonics-only condition in experiment 1B. The odd-harmonics-only condition did nevertheless yield some improvement in f_0 discrimination performance, relative to the diotic condition with all harmonics present. One possible explanation for this improvement is that listeners were able to track the frequencies of the individual harmonics, which would become more prominent with the doubling of the frequency spacing between components in any one ear, without necessarily extracting the f_0.

A similar explanation, in terms of listening to individual harmonics, may also explain the difference in f_0 DLs between the dichotic and diotic conditions observed in experiment 1A (Fig. 1). This finding contrasts with those of Bernstein and Oxenham (2003), who found, if anything, a small increase in f_0 DLs when going from diotic to dichotic conditions. This may be because the Bernstein and Oxenham (2003) randomly assigned even and odd harmonics to left and right ears on an interval-by-interval basis,[1] whereas in the current study the assignment was made on a trial-by-trial basis, with the same assignment holding for all three intervals of the trial. This may have increased listeners' ability to perform the task in the current study by tracking individual harmonics, listening selectively to one ear, rather than extracting a pitch based on the f_0. Such a strategy is most likely to have been used in the dichotic conditions with f_0's in the range of 100–175 Hz, where the effect of diotic versus dichotic presentation was observed, and where pitch discrimination was relatively poor, but peripherally resolved harmonics would have become available under dichotic

presentation. In any case, dichotic presentation in the current study yielded only a modest improvement in f_0 DLs relative to the improvement yielded by shifting the frequencies of the odd harmonics. Thus, any benefit to f_0 discrimination through increased peripheral resolvability via dichotic presentation was overshadowed by the effect of shifting the frequencies of the odd harmonics.

III. EXPERIMENT 2: MAGNITUDE OF THE ODD-HARMONIC F_0 SHIFT

A. Rationale

Experiment 1 used a 3% shift in the f_0 of the odd harmonics to perceptually segregate them from the even harmonics. This is consistent in some ways with earlier work showing that a shift of as little as 1% is sufficient for a single harmonic to be heard as a separate object against the background of the remaining complex (Moore *et al.*, 1986). On the other hand, it has also been shown that single mistuned harmonics can continue to contribute to the pitch of the overall complex at much greater mistunings (Darwin and Ciocca, 1992; Darwin *et al.*, 1995). The fact that our mistuning of 3% improved f_0 discrimination performance in a manner consistent with the perceptual segregation of the even and odd components suggests that the "harmonic sieve," outside which components fail to be combined within a single pitch estimate (Duifhuis *et al.*, 1982), is narrower than 3% for the current conditions. Experiment 2 sought to determine more accurately the bandwidth of this putative sieve by estimating the minimum odd-harmonic mistuning necessary to improve f_0 discrimination performance.

B. Methods

This experiment directly estimated the f_0 associated with the transition point between low and high f_0 DLs, rather than measuring f_0 discrimination for a range of f_0 values as was done in experiment 1. A 3I-3AFC procedure adaptively varied the f_0 while fixing the f_0 difference between the $f_{0,\mathrm{base}}$ in the two reference intervals and the higher f_0 in the target interval. The f_0 difference was fixed at 2.5% (dashed line in Fig. 1), which was selected because it fell between the small f_0 DLs (~1–2%) associated with relatively high f_0's and the large f_0 DLs (4% or greater) associated with relatively low f_0's for each listener and condition in experiment 1. The $f_{0,\mathrm{base}}$ was initially set to 250 Hz and was changed by a factor of 1.26 for the first two reversals points, 1.12 for the next two reversal points, and 1.047 for the last six reversal points. The f_0 DL transition point was estimated as the geometric mean of the f_0's at the last six reversal points.

Stimuli were presented both diotically and dichotically as in experiment 1. Six different values of Δf_0 ($f_{0,\mathrm{odd}} - f_{0,\mathrm{even}}$) were tested: 0, 1, 2, 3, 4%, and −4%. The 0 and 3% percent conditions correspond to the harmonic and f_0-shifted conditions of experiment 1. The −4% condition tested whether the effect was symmetric for both negative and positive mistunings of the odd harmonics. The same four listeners from experiment 1A participated in this experiment (Table I). Measurements were repeated four times for each listener and condition, for a total of 48 runs per listener. Each listener received at least 2 h practice before data collection began.

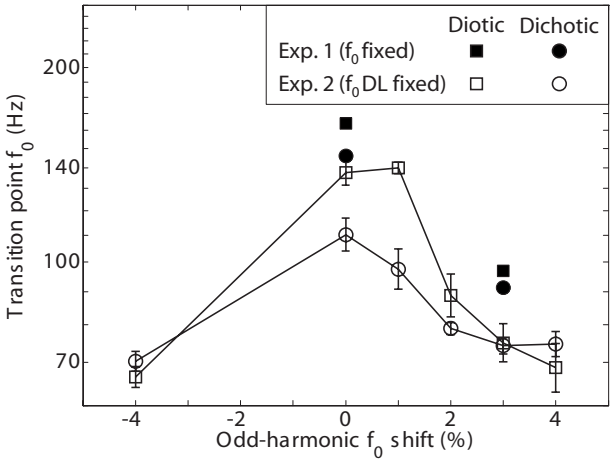


FIG. 3. Mean results of experiment 2 across four listeners. The minimum f_0 needed to yield an f_0 DL of 2.5% is plotted as a function of the magnitude of the odd-harmonic f_0 shift. Error bars indicate ± one standard error across the four listeners.

C. Results and discussion

Estimates of the f_0 DL transition point—the f_0 required to achieve an f_0 DL of 2.5%—are plotted as a function of Δf_0 for both diotic (open squares) and dichotic conditions (open circles) in Fig. 3. Three important findings are apparent in the results. First, the paradigm employed in experiment 2 yielded an estimate of the variation in the f_0 DL transition point across conditions similar to that observed in experiment 1. Estimates of the f_0 DL transition points (the f_0 required to achieve a 2.5% f_0 DL) in experiment 1 were derived by linearly interpolating between the data points in Fig. 1, and are plotted as filled squares (diotic) and circles (dichotic) in Fig. 3. Although the transition point estimates were overall 20–35% higher in experiment 1 than in experiment 2, both experiments show a shift of about a factor of 2 in the f_0 DL transition point for a 3% odd-harmonic f_0 shift in the diotic conditions, and about a 10 (experiment 1) or 20% reduction (experiment 2) in the transition point under dichotic as compared to diotic presentation in the harmonic conditions.

The second important finding is that a shift of about 2% in the f_0 associated with odd harmonics was needed to produce a shift in the f_0 transition point, although some additional benefit was observed as the size of the shift increased beyond 2%. The variation in the transition point as a function of the degree of f_0 shift was confirmed by a significant main effect of f_0 shift [$F(5.0,14.9)=42.3$, $p<0.0005$]. As in experiment 1, dichotic presentation led to improved f_0 discrimination for the harmonic (0%) condition and for small f_0 shifts (1% and perhaps 2%). This observation was supported by a significant main effect of mode of presentation [$F(1,3)=32.5$, $p<0.05$] and a significant interaction between mode of presentation and f_0 shift [$F(2.7,8.0)=5.1$, $p<0.05$]. Separate RMANOVAs confirmed that an effect of f_0 shift on the transition point was observed under both diotic

and dichotic presentation [diotic: $F(4.2,12.7)=28.1$, $p<0.0005$; dichotic: $F(5,15)=16.5$, $p<0.0005$], although the shift in the transition point was larger under diotic presentation as a result of the smaller benefit provided by dichotic presentation in the f_0-shifted conditions. The 2% odd-harmonic frequency shift required to yield the f_0 discrimination benefit is somewhat less than the 3% point at which the contribution that a mistuned harmonic makes to the overall pitch of a tone complex begins to diminish (Darwin and Ciocca, 1992; Darwin *et al.*, 1995). It may be that the simultaneous mistuning of all of the odd harmonics further encouraged their reduced contribution to the pitch percept associated with the even harmonics, such that a smaller frequency shift was needed to yield the f_0 discrimination benefit. The extent to which the f_0 discrimination benefit depends on the harmonic relationship between the frequency-shifted odd components remains an open question that could be addressed by randomizing the degree and direction of the mistuning of individual components.

Finally, the benefit to f_0 discrimination of mistuning the odd harmonics was also observed in the $\Delta f_0=-4\%$ condition where, on average, frequency components were spaced more closely together than in the 0% conditions. This suggests that the benefit obtained from frequency shifting the odd components is not due to improved harmonic resolvability because, if anything, average peripheral resolvability would decrease when the odd components are shifted lower in frequency by 4%. Experiment 3 examines more closely the effect of frequency shifting the odd components on the ability to hear out individual harmonics.

IV. EXPERIMENT 3: HEARING OUT HARMONICS

A. Rationale

Experiment 1 showed that the improvement in f_0 discrimination obtained by shifting the f_0 of the odd components was observed whether the even and odd harmonics were presented diotically or dichotically. This suggests that the observed improvement in f_0 discrimination was not related to peripheral harmonic resolvability. Nevertheless, this result does not rule out the possibility that peripheral resolvability may play a role in determining f_0 discrimination performance, as shifting the frequencies of the odd harmonics could have affected peripheral resolvability. This experiment tested this possibility directly by measuring the ability of listeners to "hear out" the frequencies of individual harmonics. An improvement in the ability to hear out harmonics as a result of odd-harmonic frequency shifting would leave open the possibility that increased peripheral resolvability may have contributed, at least in part, to the improved f_0 discrimination performance observed in experiment 1A. On the other hand, if the frequency shift does not systematically improve the ability to hear out harmonics, a role of peripheral resolvability could be ruled out as the basis for the f_0 discrimination benefit observed for f_0-shifted stimuli.

B. Methods

The method used was similar to that of Bernstein and Oxenham (2003; 2006a), whereby listeners discriminated the frequency of a pure tone presented in isolation from that of a component embedded in a tone complex. Each trial consisted of two intervals, each with a 500-ms duration, separated by 300 ms. The second interval contained a bandpass-filtered tone complex, identical to that of experiment 1, except that one harmonic (the target tone) was gated on and off in time, with three bursts of a 150 ms sinusoid, including 30 ms raised-cosine onset and offset ramps between bursts, separated by 25 ms silent gaps. The onset of the first burst and the offset of the last burst of the target tone were synchronous with the onset and offset of the remaining components in the complex, respectively. The first interval contained a single stimulus frequency (the comparison tone) gated on and off in the same manner as the target tone.

The frequency of the comparison tone (f_{comp}) was selected from a uniform distribution ranging from 1575 to 1675 Hz (near the low-frequency cutoff of the bandpass filter used to define the harmonic complexes). This eliminated the possibility that listeners could base their responses on the absolute frequency of the comparison tone alone, which would be a confounding factor if the f_{comp} were set to be higher or lower than a particular value or range of f_{targ}. For a particular trial, the f_{targ} was set to be either 4% higher or lower than the selected f_{comp} (each with probability 0.5), and the f_0 of the tone complex was set relative to the f_{targ} based on the target harmonic number. The listener was required to discriminate whether the target was higher or lower in frequency than the comparison tone. Feedback was provided following each response.

The same four stimulus conditions from experiment 1 were presented (diotic and dichotic, harmonic and f_0 shifted). Also like experiment 1, in the f_0-shifted conditions, $f_{0,odd}$ was shifted 3% higher than the $f_{0,even}$. In the dichotic conditions, the even harmonics in interval 2 and the comparison tone in interval 1 (which was always to be compared to an even target harmonic in interval 1) were presented to the right ear, while the odd harmonics were presented to the left ear. This was done, rather than randomly assigning even and odd harmonics to right and left ears as in experiment 1, so that listeners would not have to shift their attention on a trial-by-trial basis to hear out the comparison and target tones. Each run consisted of four trials for each of nine target harmonic numbers (N=4, 6, 8, 10, 12, 16, 20, 24, and 32), presented in random order, for one of the four conditions. These values of N correspond to an f_0 range of approximately 50–400 Hz for the mean f_{comp} of 1625 Hz. Twelve runs were presented for each condition, for a total of 48 runs per listener.

All stimuli were presented in the same wideband TEN background as experiment 1, which was turned on 200 ms before the start of the first interval and turned off 100 ms following the end of the second interval. Each component (before filtering, where applicable) was presented at 12.5 dB SL (adjusted for each listener). Level randomization was not used in this experiment, because overall loudness variations would not have provided a usable cue. The same four listeners (Table I) from experiments 1A and 2 participated in this experiment. Each listener received at least 2 h of practice before data collection began.

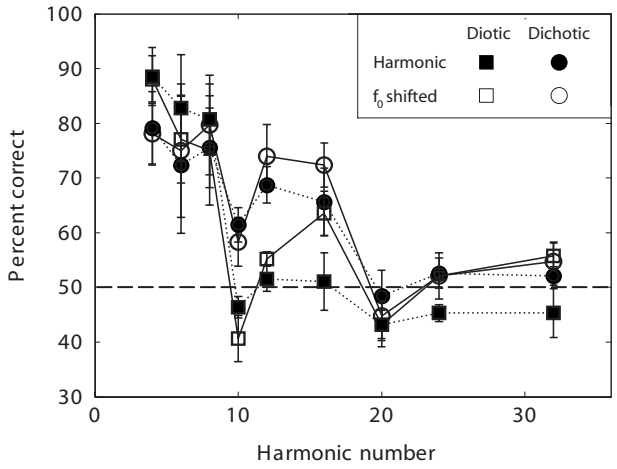


FIG. 4. Mean results of experiment 3 across four listeners. Percent correct in discriminating between a reference tone and a target frequency component of a tone complex is plotted as a function of the harmonic number of the target tone. Error bars indicate ± one standard error across the four listeners. In contrast to the f_0 DL data (Fig. 1), little or no benefit to hearing out harmonic frequencies was observed as a result of the odd-harmonic f_0 shift, while a large benefit was obtained under dichotic presentation.

C. Results

The mean percent correct scores in discriminating the frequencies of the target and comparison tones are plotted as a function of harmonic number in Fig. 4. Error bars indicate the standard error across the four listeners. The horizontal dashed line indicates chance performance. A RMANOVA with three within-listener factors (N, mode of presentation, f_0 shift) was performed to test the significance of trends observed in the data. Performance generally deteriorated from near perfect (100%) to near chance (50%) across the tested range of N in all conditions, consistent with a significant main effect of N [$F(2.5,7.4)=11.9$, $p<0.005$], as expected given previous results (Bernstein and Oxenham, 2003; 2006a). Most importantly, the results indicate that the relationship between performance and harmonic number was mainly dependent on mode of presentation (diotic or dichotic) and not the f_0 shift. Although there was no significant main effect of mode of presentation [$F(1,3)=3.1$, $p=0.18$], there was a significant interaction between mode of presentation and N [$F(8,24)=4.4$, $p<0.005$], reflecting the observation that mode of presentation affected performance mainly near the center of the range of N's presented, but not at high or low N's where ceiling and floor effects likely influenced the results. Under diotic presentation, performance dropped from about 75% or more to near chance as N increased from eight to ten. In the dichotic conditions, there was a dip in performance for $N=10$, but performance then improved again before dropping to near chance for $N=20$. This factor-of-two increase in the N at which performance dropped to chance, also observed by Bernstein and Oxenham (2003), is consistent with the idea that the ability to hear out individual harmonics is a function of peripheral resolvability, since the peripheral frequency spacing between harmonics is doubled under dichotic presentation. In contrast to the effect of mode of presentation, there was neither a significant main effect of f_0 shift [$F(1,3)=2.7$, $p=0.20$] nor any significant two- or three-way interactions between f_0 shift and the other variables (f_0 shift $\times N$: $p=0.29$; f_0 shift $\times$ mode of presentation: $p=0.93$; f_0 shift $\times N$ and mode of presentation: $p=0.51$). These results suggest that shifting the f_0 of the odd harmonics did not affect listeners' ability to hear out the individual components. It is possible that the relatively low statistical power of the ANOVA for detecting higher-order interaction effects may have rejected as nonsignificant some differences that are visually apparent in the data (e.g., the diotic $N=16$ condition). To increase the possibility of detecting the effects of, and interactions with, the f_0 shift, an additional ANOVA was performed on only those data in the diotic conditions where an f_0 shift in the odd components benefited f_0 discrimination performance in experiment 1 ($N=10$, 12 or 16). Within this subset, the main effect of N remained significant [$F(2,6)=8.2$, $p<0.05$] and the main effect of f_0 shift remained nonsignificant ($p=0.31$). The f_0 shift $\times N$ interaction just failed to reach significance ($p=0.058$), raising the possibility that listeners might have benefited from the f_0 shift in the instance of the diotic $N=16$ condition. Nevertheless, the value of N below which performance was consistently above chance was not affected by the f_0 shift.

D. Discussion

The results show that the ability to hear out the frequencies of individual harmonics depended on the harmonic number and mode of presentation (diotic or dichotic), whereas the 3% upward shift in the frequencies of the odd components did not result in a consistent or statistically significant improvement. Most importantly, performance fell to chance at the same harmonic number whether or not the odd harmonics were shifted (diotic: 10th harmonic; dichotic 20th harmonic). This finding contrasts with the f_0 discrimination results of experiments 1 and 2, where the f_0 shift greatly benefited performance, shifting by an octave the f_0 transition point between relatively poor and good f_0 discrimination performance, while dichotic presentation yielded only a limited improvement from very poor (f_0 DL ~10%) to less poor (f_0 DL ~5%) performance generally associated with unresolved harmonics. The possibility that the f_0 shift improved the ability to hear out harmonics under limited circumstances (e.g., for $N=16$) cannot be completely ruled out. Nevertheless, there was no clear evidence of a systematic improvement in the ability to hear out the frequencies of individual harmonics as a result the harmonic f_0 shift. Therefore, it is unlikely that the substantial improvement in f_0 discrimination across a range of f_0's that was observed in experiment 1 could be explained in terms of an improved ability to hear out individual harmonics. Instead, these results generally support the conclusions of experiment 1 that f_0 discrimination can be improved by perceptual segregation mechanisms (i.e., mistuning the odd harmonics), but not by the increased peripheral resolvability of harmonics (i.e., dichotic presentation).

Nonmonotonicities were observed in all four conditions of this experiment, whereby the performance functions each showed a local minima at $N=10$. Bernstein and Oxenham (2003; 2006a) observed similar nonmonotonicities near N

=10. This result could reflect the phenomenon of "unmasking" (Hartmann and Goupell, 2006), whereby the frequency component just above the target harmonic becomes "separately audible" during the silent intervals of the gated harmonic. Discrimination judgments based on the frequency of this salient "unmasked" harmonic rather than that of the gated target harmonic would yield diminished performance. However, it is not clear why the unmasking phenomenon would only have this effect for $N=10$ and not other harmonics. Further experiments involving pitch matching to the gated harmonic (e.g., Hartmann and Goupell, 2006), and not simply pitch discrimination, may shed light on this result.

V. EXPERIMENT 4: PITCH MATCHES

A. Rationale

The main finding from experiments 1 and 2 was that shifting the f_0 of the odd harmonics, with respect to that of the even harmonics, resulted in improved f_0 discrimination performance. The pattern of results can be explained by assuming (1) that shifting the frequencies of the odd components decreased their contribution to the perceived pitch associated with the even components, and (2) that the resulting octave pitch shift yielded improved f_0 discrimination performance because performance is based on the *perceived* harmonic number—that is, the ratio between the absolute frequency and the f_0 associated with the perceived pitch of a stimulus. This experiment tested whether the f_0 shift of the odd harmonics did indeed produce an octave pitch shift in those conditions that yielded improved f_0 discrimination performance. A pitch-matching paradigm was used to determine the pitch perceived under the various conditions of experiment 1. Experiment 4A examined pitch matches to the diotic and dichotic f_0-shifted and the dichotic harmonic stimuli of experiment 1A. Experiment 4B examined pitch matches for the stimuli consisting of only odd harmonics from experiment 1B.

B. Methods

In experiment 4A, four listeners (Table I) performed pitch matches by comparing the pitch of a diotic harmonic tone complex (assumed to yield a pitch at the f_0, regardless of harmonic number) with the pitch of a tone complex from one of the other three altered-stimulus conditions that were presented in experiment 1 (diotic f_0 shifted, dichotic harmonic, or dichotic f_0 shifted). The purpose of this experiment was to determine the perceived pitch associated with each altered stimulus. To address this question, the f_0 of a particular altered stimulus was held fixed as the reference (the altered-reference condition, AREF) while the listener adjusted the f_0 of the diotic harmonic complex to match the pitch. To test whether the pitch matches would depend on which of the two stimulus f_0's was controlled by the listener, a control condition was also included, whereby the f_0 of the diotic tone complex was held fixed while the listener adjusted the f_0 of the altered stimulus (the altered-comparison condition, ACOMP).

In each pitch-matching run, two tone complexes were presented, each with a 300-ms duration, separated by 500 ms. The tone complexes were constructed in an identical manner to those described in experiment 1, including the background TEN, except that level and f_0 roving were not applied. The tone complex for which the listener had control over the f_0 was presented first (comparison stimulus), followed by the tone complex that was held constant throughout the run (reference stimulus). For the reference stimulus, the starting phase of each component was randomly selected from a uniform distribution, and then held fixed throughout the run. For the comparison stimulus, the starting phases were randomly selected each time the comparison f_0 ($f_{0,\mathrm{comp}}$) was adjusted. Following each presentation of the two sequential complexes, the listener had eight choices, selected by virtual buttons displayed on a computer monitor, and selected via mouse click: (1–6) increase or decrease the $f_{0,\mathrm{comp}}$ by a large, medium, or small amount, (7) hear the same two stimuli again without manipulating the $f_{0,\mathrm{comp}}$, or (8) indicate that the match was satisfactory. The step sizes for the large, medium, or small f_0 adjustments were 4, 1, and $\frac{1}{4}$ semitones, respectively. The smallest step size in the pitch matching procedure (corresponding to about 1.5%) would not be expected to be sensitive enough to detect small shifts in the perceived pitch on the order of 1–3% that have been observed in other studies of the effect of frequency shifting individual components on the perceived pitch (e.g., Moore *et al.*, 1985; Darwin and Ciocca, 1992). However, this experiment was mainly concerned with testing the hypotheses that the shift in the f_0 DL transition point by a factor of about 2, observed in experiment 1, is related to an octave shift in the perceived pitch.

In pilot runs, it was found that listeners almost always matched an altered tone complex (i.e., f_0 shifted and/or dichotic) with a diotic harmonic tone complex that was near either the f_0 or twice the f_0 ($2f_0$) of the altered complex. Therefore, the range of starting f_0's was specified to be symmetrical (on a log scale) around this zero- to one-octave range of matches. For AREF runs, the starting $f_{0,\mathrm{comp}}$ was randomly selected from a range of −0.25 to +1.25 octaves relative to the reference f_0 ($f_{0,\mathrm{ref}}$). The range of possible $f_{0,\mathrm{comp}}$ values was limited to the range of −1 to +2 octaves relative to the $f_{0,\mathrm{ref}}$. If a listener attempted to increase or decrease the $f_{0,\mathrm{comp}}$ outside of this range, the $f_{0,\mathrm{comp}}$ would simply stay at its previous value for the next stimulus presentation. For ACOMP runs, the starting $f_{0,\mathrm{comp}}$ was randomly selected from a range of −1.25 to +0.25 octaves and the range of $f_{0,\mathrm{comp}}$ was limited to −2 to +1 octaves relative to the $f_{0,\mathrm{ref}}$. An analysis of the results (not shown) indicated that there was no systematic relationship between the starting value of $f_{0,\mathrm{comp}}$ and the pitch match.

For each of the three altered tone complex conditions (diotic shifted, dichotic shifted, and dichotic harmonic), each listener completed ten AREF pitch matches with the f_0 of the altered complex held fixed at each of the seven f_0's that were tested in experiment 1. Each listener also completed ten ACOMP pitch matches each with the f_0 of the diotic harmonic complex held fixed at 100, 125, 150, 175, and 200 Hz, for a total of 360 pitch matches. The 50- and 75-Hz f_0's were not included in the latter pitch matches because listeners were generally unable to provide a reliable match of altered

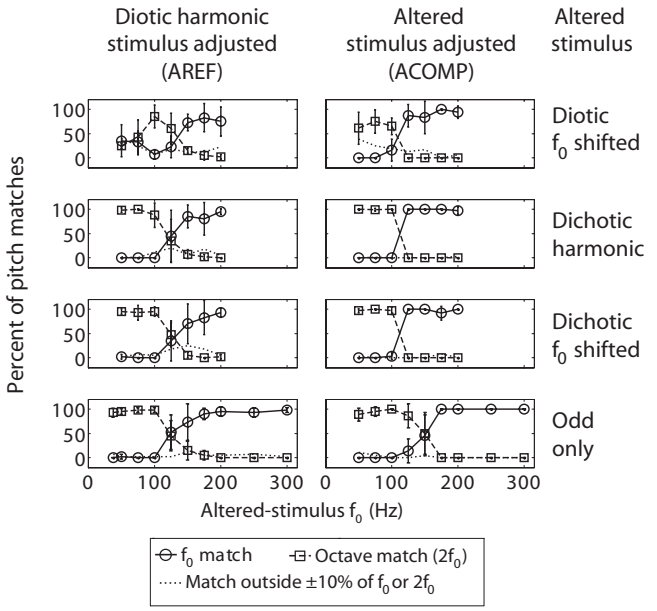


FIG. 5. Mean results of experiment 4 across four listeners. Left column: Summary of pitch matches for the AREF conditions, where the f_0 of the altered tone complex (dichotic, f_0-shifted or odd-only) is held constant and the listener adjusts the f_0 of the diotic harmonic complex. The proportion of pitch matches that were within $\pm 10\%$ of the $f_{0,\mathrm{ref}}$ (circles) or twice the $f_{0,\mathrm{ref}}$ (squares) are plotted as a function of the $f_{0,\mathrm{ref}}$. Right column: Summary of pitch matches for the ACOMP condition, where the f_0 of the diotic harmonic complex is held fixed and the listeners adjust the f_0 of the altered stimulus to match. The proportion of pitches that were within $\pm 10\%$ of the $f_{0,\mathrm{ref}}$ (circles) or half the $f_{0,\mathrm{ref}}$ (squares) are plotted as a function of the $f_{0,\mathrm{ref}}$. For both columns, the dotted line indicates the proportion of matches that did not fall within $\pm 10\%$ of the $f_{0,\mathrm{ref}}$ or within $\pm 10\%$ an octave above (left panel) or below (right panel) the $f_{0,\mathrm{ref}}$. Error bars indicate $\pm$ one standard error across the four listeners.

tone complex to these low-pitched diotic harmonic stimuli. For conditions involving dichotic complexes, even and odd components were assigned to the left and right ear, respectively, for five of the ten pitch-match runs, and *vice versa* for the remaining five runs. Before the testing phase, each listener completed one practice pitch match for each condition (for a total of 36 practice runs). An additional experiment (4B) determined the perceived pitch associated with the odd harmonics only. Four listeners (Table I) completed the pitch matching task by comparing the pitch associated with a complex containing only odd harmonics to a complex containing all harmonics for the same ten f_0's that were tested in experiment 1B. Each listener received at least 20 min practice before data collection began.

C. Results

The AREF data (left column of Fig. 5) indicate the perceived pitch of each altered stimulus. Across all listeners and conditions, the overwhelming majority of pitch matches were close to the f_0 of the reference stimulus or its octaves. In the AREF conditions, the diotic harmonic stimulus was matched to within 10% of either the $f_{0,\mathrm{ref}}$ or $2f_{0,\mathrm{ref}}$ on 94% of the trials across listeners. In the ACOMP conditions, the altered stimulus was matched to within 10% of either 0.5 $f_{0,\mathrm{ref}}$ or $f_{0,\mathrm{ref}}$ on 91% of the trials. The left column of Fig. 5 plots the percentage of trials for which each of these outcomes occurred as a function of the $f_{0,\mathrm{ref}}$ (the ACOMP data shown in the right column are discussed below). Each row represents one of the four altered-stimulus conditions: Diotic f_0-shifted, dichotic harmonic, dichotic f_0-shifted, and odd only. Circles and squares show the proportion of responses for which the perceived pitch of the altered stimulus was within 10% of the f_0 or within 10% of twice the f_0, respectively. The dashed lines indicate the proportion of pitch matches that did not fall within $\pm 10\%$ of the f_0 or $2f_0$.

Pitch matches at each f_0 can be grouped into four categories. The first two categories are a clear match at the f_0 or its octave, indicated by circles or squares near 100%, respectively, suggesting a clear and unambiguous pitch percept. The third category is a bimodal distribution of pitch matches, where matches were equally apportioned between the f_0 and its octave, indicated by circles and squares both near 50% (i.e., the 125-Hz f_0 in all four panels), suggesting either a bistable percept, where sometimes one pitch was heard, sometimes another, or an ambiguous pitch that could not be clearly assigned to either octave. The large error bars in these cases indicate that some subjects consistently matched to the f_0, while others matched to the octave. The fourth and final category is a broad distribution of matches, with a roughly equal number of matches near the f_0, its octave, or outside of the $\pm 10\%$ ranges surrounding these values (i.e., the 50- and 75-Hz diotic f_0-shifted conditions), suggesting an ambiguous and unstable pitch percept.

Overall, these plots indicate that each stimulus alteration (dichotic presentation, f_0 shift and odd-harmonics only) had a similar effect on the perceived pitch. The pitch was roughly equal to the f_0 for altered-stimulus f_0's above 125 Hz, and equal to $2f_0$ for altered-stimulus f_0's below 125 Hz, with a bistable pitch at 125 Hz (and 150 Hz in the odd-only condition). The only exceptions to this trend were for low f_0's of 50 or 75 Hz in the diotic f_0-shifted condition, where the pitch was ambiguous, often producing matches outside of the $\pm 10\%$ range surrounding the f_0 and $2f_0$. The ambiguous pitch in these diotic f_0-shifted conditions may reflect the complex beat patterns that result from peripheral interactions between the even and frequency-shifted odd components.

The ACOMP data are plotted in the right column of Fig. 5. For direct comparison with the AREF data, these data are also plotted as a function of the f_0 of the altered stimulus, which in this case was the dependent variable over which the listener had control. For each altered-stimulus type, the matched f_0's were grouped into 25-Hz-wide bins with midpoints ranging from 50 to 200 Hz (or 50 to 300 Hz in the odd-only conditions). The percent of trials (averaged across listeners) for which the matched altered-stimulus f_0's that fell in each bin were within $\pm 10\%$ of the fixed diotic harmonic reference or its octave are plotted as circles and squares, respectively. For some combinations of f_0 and condition (especially f_0's ≥ 100 Hz), at least one of the listeners had no pitch matches that fell within a particular bin. In such cases, the proportion of f_0 or octave matched trials were averaged across the remaining listeners.

Overall, the ACOMP conditions yielded very similar results to the AREF conditions, with only two small differences. First, the crossover from a match at the f_0 to a match

at the octave occurred at a slightly different f_0 DL in the ACOMP conditions (higher than the AREF crossover in the odd-only conditions, lower in the other conditions). Second, the matches at the lowest f_0's in the diotic f_0-shifted conditions were less ambiguous in the ACOMP than in the AREF case. One possible explanation for these slight differences is that listeners had control over the altered-stimulus f_0 in the ACOMP conditions, and therefore could have steered away from more ambiguous pitches associated with these stimuli.

D. Discussion

In experiment 1A, the stimuli that showed the most benefit from the f_0 shift were those with a 100- or 125-Hz f_0, and to some extent a 150-Hz f_0 (Fig. 1). The upper left panel of Fig. 5 shows that for the 100- and 125-Hz f_0's, the diotic f_0-shifted stimuli yielded an octave shift in the perceived pitch. The octave shift in the perceived pitch is consistent with a pitch based on the even components alone, and suggests that for these stimuli, f_0 discrimination performance is determined by the harmonic number of the perceived pitch of the stimulus.

For the dichotic stimuli, the pitch matches were the same in the harmonic and f_0-shifted conditions (second and third rows of Fig. 5). For the dichotic harmonic stimuli, the doubling of the f_0 match relative to the diotic harmonic condition was not associated with an improvement in a shift in the f_0 DL transition point. In this case, poor pitch discrimination performance might be expected if the pitch were extracted from the envelope repetition rate, which would be shifted by an octave relative to the diotic case as result of the doubling of the peripheral spacing between adjacent components in each ear. Although the octave shift in the 100 and 125 Hz dichotic f_0-shifted conditions (Fig. 5, third row), where an f_0 discrimination improvement was observed in experiment 1A, might also reflect a doubling in the envelope repetition rate, it would be surprising for a pitch based on the envelope repetition rate to be as discriminable as the pitch associated with resolved harmonics, as was observed for the 100–150 Hz f_0-shifted dichotic conditions in experiment 1A. This argues in favor of the idea that the octave shift reflects a pitch percept associated with the even components alone.

Taken together, the results of the pitch matching and f_0 discrimination data argue that f_0 discrimination performance is determined by the harmonic number of the perceived pitch of the stimulus, except in cases where the pitch is extracted from the envelope repetition rate, in which case discrimination is always poor.

VI. GENERAL DISCUSSION

A. The role of the auditory periphery

The main goal of this study was to determine the extent to which peripheral and/or central harmonic resolvability governs performance in f_0 discrimination tasks. In terms of the three hypotheses laid out in the introduction: (1) harmonic resolvability, as defined by a "central spectrum" combining the spectra from the two ears, does not appear to limit performance; (2) peripheral (monaural) resolvability also seems not to limit performance; and (3) good f_0 discrimination seems not to depend directly on harmonic resolvability at all, but instead on the harmonic number associated with the perceived f_0. These results therefore appear to provide strong evidence against models of pitch perception that depend solely on spectrally resolved harmonics for good performance (e.g., Goldstein, 1973; Wightman, 1973; Terhardt, 1974).

Taken in isolation, these results might suggest little or no relationship between peripheral auditory filtering and f_0 discrimination. This conclusion would, however, contradict our earlier findings of a strong relationship between auditory filter bandwidths and f_0 discrimination, both as a function of overall level in normal-hearing listeners (Bernstein and Oxenham, 2006a), and as a function of degree of hearing loss (and filter widening) in hearing-impaired listeners (Bernstein and Oxenham, 2006b). Instead, a parsimonious interpretation of the available results is that f_0 discrimination depends on auditory filter bandwidth, but not on harmonic resolvability *per se*. Of course, in most everyday situations the two measures co-vary, as wider auditory filters imply poorer harmonic resolvability. It is only through the technique of mistuning harmonics to induce perceptual segregation (present study), and of presenting harmonics to opposite ears (Bernstein and Oxenham, 2003), that we were able to dissociate them.

We are aware of only one model that explicitly dissociates auditory filter bandwidth from harmonic resolvability, as our data suggest should be the case. As mentioned in the introduction, the model of de Cheveigné and Pressnitzer (2006) is explicitly based on filter bandwidths, but does not depend on harmonic resolution and is based instead on a variant of the temporal autocorrelation function (Licklider, 1951; Meddis and O'Mard, 1997). However, that model, in its current form, is not sufficiently developed to provide quantitative predictions for our data. The following section uses another variant of the autocorrelation function to test whether such an approach can, in principle, account for the perceptual effects of mistuning odd and even harmonics from each other.

B. Autocorrelation model

Autocorrelation (AC) models of pitch perception (Licklider, 1951; Meddis and Hewitt, 1991a, b; Meddis and O'Mard, 1997) account for the human ability to extract the missing f_0 based on periodic temporal information in auditory nerve fiber (ANF) responses. The AC model proposed by Bernstein and Oxenham (2005) is a modification of that of Meddis and O'Mard (1997), in which individual AC functions are first calculated in a population of simulated ANFs with characteristic frequencies (CFs) distributed across the tonotopic range of the cochlear partition, and then summed to produce a single f_0 estimate. Like the AC model of de Cheveigné and Pressnitzer (2006), that of Bernstein and Oxenham (2005) contains a CF-dependent limitation on the range of lags for which the AC can be computed—an essential ingredient in accounting for the deterioration in f_0 discrimination performance with increasing harmonic number,

independent of absolute frequency. These models differ in that this lag-range limitation is directly related to the characteristics of peripheral auditory filters in the case of de Cheveigné and Pressnitzer (2006), but is applied in an *ad hoc* manner by Bernstein and Oxenham (2005), such that the latter model is unlikely to account for the effects of broadened auditory filters on f_0 discrimination observed by Bernstein and Oxenham (2006a, b). This difference should be largely inconsequential for the current experiments, where stimuli were presented at a fixed level in normal-hearing listeners, such that auditory filter characteristics were unlikely to vary across conditions.

Meddis and O'Mard (1997) showed that an AC model of pitch perception was able to account for the decreased influence of a mistuned harmonic on the overall pitch percept associated with a harmonic complex as the degree of mistuning increases (Darwin *et al.*, 1994). Given this result, we hypothesized that the modified AC model of Bernstein and Oxenham (2005) might account for the improved f_0 discrimination resulting from the 3% odd-harmonic mistuning in the current study, as follows. As the odd harmonics become mistuned, their contribution to the pitch associated with the even harmonics will be reduced. Since the predicted f_0 discrimination performance of the model depends on the harmonic number associated with the perceived f_0 (i.e., the ratio between the first peak in the autocorrelation function and the absolute frequency region of the stimulus), a doubling of the perceived pitch might yield improved predicted performance. To test this hypothesis, the diotic harmonic and f_0-shifted stimuli of experiment 1A were presented to the modified AC model of Bernstein and Oxenham (2005).

The modified AC model consists of an outer/middle ear bandpass filter, a gammatone filterbank (Patterson *et al.*, 1992) consisting of 40 channels with characteristic frequencies (CFs) logarithmically spaced between 1.5 and 5 kHz to simulate basilar-membrane filtering, followed by a model of inner hair cell and auditory nerve processing (Sumner *et al.*, 2002). An autocorrelation function in each channel was calculated for the binary stochastic spike train generated in response to each stimulus. The periodicity-range limitation was then applied by weighting the calculated autocorrelation in each channel relative to the channel's CF using the parameter values given by Bernstein and Oxenham (2005) that best fit the psychoacoustic f_0 discrimination data described in that study.[2] A summary autocorrelation function (SACF) was produced by adding the weighted autocorrelation functions across channels. This was repeated 100 times for each stimulus. An optimal-detector d' metric (Van Tress, 2001) was then generated for pairs of stimuli differing in f_0 for 30 log-spaced values of Δf_0 ranging from 0.5% to 30% of the f_0. The f_0 DL estimate was defined as the minimum Δf_0 to yield a value of d' greater than some fixed threshold value. The only deviations from the modeling procedure described in Bernstein and Oxenham (2005) were (1) the threshold d' was manipulated to allow a better fit to the data, and (2) d' was defined as exceeding threshold if it remained above threshold for three consecutive values of Δf_0 (instead of the requirement that d' exceed threshold for all larger values of Δf_0 tested).

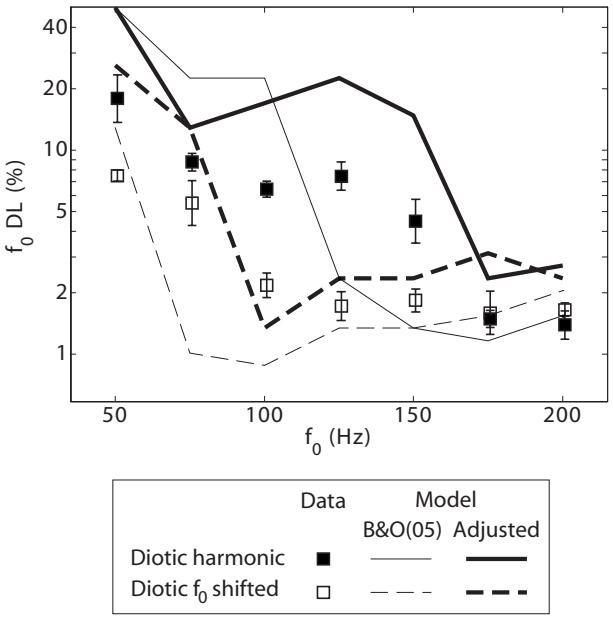


FIG. 6. Predictions of the Bernstein and Oxenham (2005) autocorrelation model of f_0 discrimination performance for the diotic harmonic (solid lines) and f_0-shifted (dashed lines) tone complexes presented in experiment 1A. Thin lines show model predictions with model parameters as specified by Bernstein and Oxenham (2005). Thick lines show model predictions with lag-range limitation parameters adjusted to better account for the locus of the transition from large to small f_0 DLs. Diotic data from experiment 1 (Fig. 1) are replotted as closed (harmonic) and open (f_0-shifted) squares.

Figure 6 replots the data from the diotic conditions of experiment 1 (symbols) and the results of the model simulations (lines). The thin solid and dashed lines show model predictions for the harmonic and f_0-shifted conditions, respectively, based on the set of model parameters defined by Bernstein and Oxenham (2005), except that the threshold d' was changed to 6×10^4 (instead of 7.91×10^4). These very large values of d' emerge because of the large number of simulated ANFs and autocorrelation lag points providing information about the stimulus f_0, and because no attempt was made to add further "internal noise" to limit performance. The model successfully accounted for the decrease in f_0 DLs with increasing f_0 seen in the psychoacoustic data of experiment 1A. This correct behavior of the model as a function of f_0 was expected given the modeling results of Bernstein and Oxenham (2005), and is a result of the CF-dependent lag window modification applied to the individual channel autocorrelation calculations. More importantly for the current study, the model also accounted for the improvement in f_0 discrimination performance in the f_0-shifted condition, with the f_0 DL transition shifting toward lower f_0's by approximately a factor of 2. The reason for the successful prediction of the f_0-shift benefit can be found in the SACFs that underlie the model's f_0 DL estimates. Mean SACFs across 100 stimulus presentation are shown in Fig. 7 for harmonic and f_0-shifted stimuli at three different f_0's (50, 100, and 200 Hz). For each f_0, the odd-harmonic f_0 shift introduces an additional SACF peak at half the lag (double the periodicity) of the first peak of the harmonic SACF. In the harmonic condition, peaks at this lag in the individual channel AC functions associated with the even components tend to de-

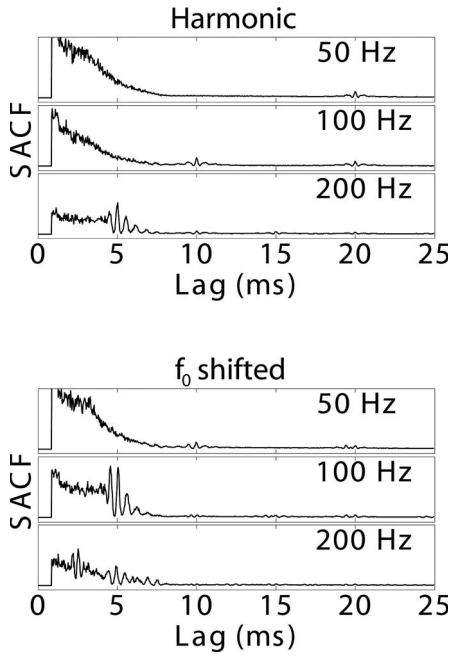


FIG. 7. Mean SACFs across 100 stimulus presentations produced by the Bernstein and Oxenham (2005) autocorrelation model in response to diotic tone complexes that were presented to listeners in experiment 1. The effect of the odd-harmonic f_0 shift (lower panel) was to produce an additional SACF peak at half the lag of the first peak for the corresponding harmonic condition (upper panel). This additional peak occurs at a more favorable lag relative to the CF-dependent lag-range limitation, yielding the f_0 discrimination benefit observed in Fig. 6.

structively interfere with the AC functions associated with the odd components. With the f_0 of the odd components shifted, the destructive interference is removed and the additional peak appears. This additional peak yields an SACF function similar to that observed for the harmonic stimulus at $2f_0$ (compare the 50- and 100-Hz f_0-shifted conditions to the 100- and 200-Hz harmonic conditions, respectively). The additional SACF peak falls at a more favorable lag relative to the autocorrelation lag weighting function, yielding the f_0 discrimination benefit.

The model was not entirely successful at accounting for the diotic f_0 discrimination data. First, the model predicted a larger difference between the f_0 DLs for low and high f_0's than was observed experimentally, a failing that was also noted by Bernstein and Oxenham (2005). Second, the locus of the transition point from large to small f_0 DLs occurs at a lower range of f_0's in the model results (100–125 Hz and 50–75 Hz for the harmonic and f_0-shifted conditions, respectively) than in the experimental data (125–175 Hz and 75–100 Hz, respectively). The locus of this transition in the model more closely matched that of the data when the lower cutoff of the lag-range limitation was adjusted to a harmonic number of 8.0 (instead of 10.8) and of the d' threshold was set to 4.2×10^4 (thick solid and dashed lines in Fig. 6). However, this change in parameters would also be likely to shift the f_0 DL transition toward a lower harmonic number in the modeling results of Bernstein and Oxenham (2005), thereby yielding a poorer fit to the data in that study. Some further adjustment of the lag-window parameters may be required to fit a broader range of f_0 discrimination data than those presented in the current study and that of Bernstein and Oxenham (2005).

The place-dependent autocorrelation models of Bernstein and Oxenham (2005) and de Cheveigné and Pressnitzer (2006) are in principle consistent with the relative lack of benefit to f_0 discrimination performance provided by dichotic presentation of even and odd harmonics observed in experiment 1A and by Bernstein and Oxenham (2003), because they do not depend on peripheral resolvability to account for harmonic-number dependence of f_0 DLs. An initial attempt was made to test the Bernstein and Oxenham (2005) model for the dichotic stimuli presented in experiment 1A. Discrimination predictions were generated using the same procedure described above, except that SACFs were calculated by summing the SACFs from the two ears. This simulation did not produce satisfactory results, predicting a large deficit in performance under dichotic presentation (results not shown). Further work is needed to determine whether this is a basic failing of the model. It may be that a different method of combining binaural information in the model would yield more satisfactory results.

C. Concurrent source segregation

Although the predictions of the modified AC model are generally consistent with the diotic f_0 discrimination data, they do not generally agree with the pitch matching data of experiment 4 (Fig. 5). The model predicts a doubling of the perceived pitch with the introduction of the f_0 shift across all f_0's, as evidenced by the appearance of the additional SACF peak at half the lag. In contrast, an octave shift in the perceived pitch was only observed experimentally at 100 and 125 Hz (Fig. 5, upper left panel). This discrepancy may be reconciled if for higher f_0's listeners were segregating the even and odd harmonics into separate objects, each with its own pitch, rather than extracting a single pitch from the stimulus as a whole. While segregation is more likely to have occurred in the dichotic conditions where f_0 and ear of presentation were both available as segregation cues, listeners may have been able to segregate even and odd harmonics in at least some of the diotic f_0-shifted conditions, especially for f_0's of 175 Hz and above where resolved harmonics were most likely available. This raises an interesting paradox at 150 Hz (lowest harmonic=10) where resolved harmonics were not available (Fig. 4), but perceptual segregation of two distinct percepts is needed to reconcile the data and model results. Indeed, the results of Micheyl *et al.* (2006) suggest that listeners may be able to hear out the pitch of a complex tone with low-order harmonics in the presence of a second complex tone for a large f_0 differences (seven-semitones or about a 50%), even when the two complexes would not yield any resolved harmonics when combined. Perhaps the perceptual segregation of the mistuned odd harmonics from the even harmonics is facilitated by the fact that many odd harmonics are mistuned simultaneously. In conditions close to the limits of resolvability (i.e. the lowest harmonic number =10), this effect may be sufficient to yield two distinct pitch percepts.

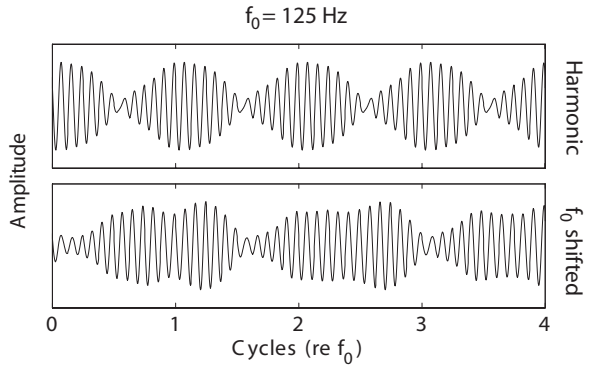


FIG. 8. Output of a single gammatone filter with CF=1500 Hz in response to a 125 Hz diotic harmonic tone complex (upper panel) and an f_0-shifted tone complex (lower panel).

D. Temporal fine structure

Moore *et al.* (2006) have argued that the deterioration in f_0 discrimination performance with increasing harmonic number may reflect a reduction in the usefulness of temporal fine structure information, rather than just a progressive reduction in peripheral harmonic resolvability. While the lack of a limiting role of peripheral resolvability observed in the current study is generally consistent with the view, the question remains as to whether the odd-harmonic f_0 shift improved the availability of pitch cues in the temporal fine structure. Moore (1982) suggested that for stimuli where harmonics are unresolved and therefore interact within individual auditory filters, the pitch could be extracted from the fine-structure peak located near the envelope peak. According to this argument, the presence of multiple fine-structure peaks of similar amplitude occurring near the envelope peak would yield a less precise estimate of the pitch than a waveform with only one high-amplitude fine-structure peak per period. Figure 8 shows a snapshot of the output of a single fourth-order gammatone filter (Patterson *et al.*, 1992), centered at 1500 Hz, in response to 125-Hz random-phase harmonic (upper panel) and f_0-shifted (lower panel) tone complexes. The filter at 1500 Hz was chosen because it represents the low-frequency edge of the stimulus bandpass filter, where the lowest-order harmonics (that generally yield the best discrimination performance) are present. While the odd-harmonic f_0 shift greatly benefited performance for the 125 Hz condition in experiment 1A, if anything, a greater number of prominent fine-structure peaks appear near the peaks in the envelope for f_0-shifted than for the harmonic stimulus in Fig. 8, which according to Moore (1982) should lead to a less discriminable pitch percept. Therefore, although an explanation for the f_0-shift benefit in terms of the temporal fine structure argument of Moore *et al.* (2006) cannot be ruled out, these plots do not appear to be consistent with such an explanation.

VII. SUMMARY AND CONCLUSIONS

Fundamental frequency discrimination was measured for bandpass-filtered harmonic complexes as a function of f_0. In line with earlier studies, f_0 DLs increased (worsened) substantially when harmonics below about the 10th were no longer present. However, when the odd harmonics were mistuned by a constant percentage, performance improved to the extent that the results were the same as when only even harmonics were present. Similar patterns of results were observed whether the odd and even harmonics were presented to the same or different ears (experiment 1). The amount of mistuning necessary to eliminate the perceptual interference from the odd harmonics was about 3%, although 2% was sufficient to observe some release (experiment 2). Although mistuning the odd harmonics dramatically improved f_0 DLs, it had no reliable effect on the ability of listeners to hear out individual harmonics, suggesting that the mistuning did not systematically improve the resolvability of the harmonics (experiment 3). In the f_0 region over which harmonic mistuning improved performance, the mistuning typically led to a doubling in the perceived f_0, in line with expectations from a pitch based on the even harmonics only (experiment 4). Taken together with previous studies, these results indicate that peripherally resolved harmonics are in themselves neither necessary nor sufficient to support accurate pitch perception.

ACKNOWLEDGMENTS

This work was supported by NIH Grant No. R01 DC 05216, and was carried out while both authors were at the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA. The authors thank Christophe Micheyl, Van Summers, three anonymous reviewers, and the associate editor, Richard Freyman, for providing helpful comments on an earlier version of the manuscript. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or Department of Defense.

[1]Bernstein and Oxenham (2003) stated that the assignment of even and odd harmonics to the left and right ears was performed on a trial-by-trial basis, implying that each group of harmonics was presented to the same ear in each of the three intervals of the trial. In fact, the assignment of harmonics to ears was randomized on each *interval*. Thus, tracking individual peaks in the excitation pattern in any one ear would not have been a reliable strategy.

[2]There was an error in the AC weighting function described by Bernstein and Oxenham (2005). Equation (7), describing the slope (m) of one section of the function, is missing a constant in the first term of the numerator, and should read as follows:

$$m = \frac{CF^2/CF_0 - A + (A/l_0)[(N_C + N_\Delta)/CF]}{N_\Delta/CF}. \quad (1)$$

See Bernstein and Oxenham (2005) for definitions of the terms in this equation.

Arehart, K. H., and Burns, E. M. (**1999**). "A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss," J. Acoust. Soc. Am. **106**, 993–997.

Beerends, J. G., and Houtsma, A. J. M. (**1986**). "Pitch identification of simultaneous dichotic two-tone complexes," J. Acoust. Soc. Am. **80**, 1048–1055.

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G. W., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on

fundamental frequency discrimination," J. Acoust. Soc. Am. **117**, 3816–3831.
Bernstein, J. G. W., and Oxenham, A. J. (**2006a**). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," J. Acoust. Soc. Am. **120**, 3916–3928.
Bernstein, J. G. W., and Oxenham, A. J. (**2006b**). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," J. Acoust. Soc. Am. **120**, 3929–3945.
Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.
Darwin, C. J., Ciocca, V., and Sandell, G. J. (**1994**). "Effects of frequency and amplitude modulation on the pitch of a complex tone with a mistuned harmonic," J. Acoust. Soc. Am. **95**, 2631–2636.
Darwin, C. J., Hukin, R. W., and al-Khatib, B. Y. (**1995**). "Grouping in pitch perception: Evidence for sequential constraints," J. Acoust. Soc. Am. **98**, 880–885.
de Cheveigné, A., and Pressnitzer, D. (**2006**). "The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction," J. Acoust. Soc. Am. **119**, 3908–3918.
Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am. **71**, 1568–1580.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
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 Goupell, M. J. (**2006**). "Enhancing and unmasking the harmonics of a complex tone," J. Acoust. Soc. Am. **120**, 2142–2157.
Houtsma, A. J. M., and Goldstein, J. L. (**1972**). "The central origin of the pitch of complex tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. **51**, 520–529.
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.
Huynh, H., and Feldt, L. S. (**1976**). "Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs," J. Educ. Stat. **1**, 69–82.
Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–133.
Meddis, R., and Hewitt, M. (**1991a**). "Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I: Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.
Meddis, R., and Hewitt, M. (**1991b**). "Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. II: Phase sensitivity," J. Acoust. Soc. Am. **89**, 2882–2894.
Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.
Micheyl, C., Bernstein, J. G. W., and Oxenham, A. J. (**2006**). "Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise," J. Acoust. Soc. Am. **120**, 1493–1505.
Moore, B. C. J. (**1982**). *An Introduction to the Psychology of Hearing*, 2nd ed. (Academic, London).
Moore, B. C. J., Glasberg, B. R., Flanagan, H., 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., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.
Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**, 205–224.
Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.
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.
Summer, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). "A revised model of the inner-hair cell and auditory-nerve complex," J. Acoust. Soc. Am. **111**, 2178–2188.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Van Tress, H. L. (**2001**). *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York).
Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.
Zurek, P. M. (**1979**). "Measurements of binaural echo suppression," J. Acoust. Soc. Am. **66**, 1750–1757.

The effect of the base line response on self-adjustments of hearing aid gain

Gitte Keidser,[a] Harvey Dillon, and Elizabeth Convery
National Acoustic Laboratories, 126 Greville Street, Chatswood, New South Wales 2067, Australia

(Received 10 October 2007; revised 29 May 2008; accepted 2 June 2008)

Self-adjustments of variable hearing aid parameters are essential for trainable hearing aids to provide customized amplification for different listening environments. Prompted by a finding of Dreschler *et al.* [Ear Hear. **29**, 214–227 (2008)], this study investigates the effect of the base line (starting) response on self-adjustments of gain in different frequency bands. In a laboratory test, 24 hearing-impaired listeners adjusted the bass, treble, and overall gain to reach their preferred responses from two different base line responses for 12 different listening situations. The adjustments were repeated five times using the preferred response after each adjustment as base line response for the next adjustment. Half of the listeners further compared three different response shapes, within the range of preferred responses, pairwise ten times for preferential and perceptual discrimination. The results revealed that base line response biases were more pronounced at low frequencies and for listeners with a flat hearing loss configuration. While 83% of listeners reliably discriminated between the average selected biased responses, only 25% demonstrated reliable preferences for one response over the other. Listeners who showed preferential discrimination ability were those who were less biased by the base line response. The clinical implication is that self-adjustments should begin from an appropriately prescribed starting response.
 [DOI: 10.1121/1.2951500]

PACS number(s): 43.66.Ts, 43.66.Sr, 43.66.Yw [BCM] Pages: 1668–1681

I. INTRODUCTION

The National Acoustic Laboratories (NAL) has been involved in the investigation of the concept of a trainable hearing aid for the past decade (Dillon *et al.*, 2006; Keidser *et al.*, 2007; Zakis *et al.*, 2007; Dreschler *et al.*, 2008). A trainable hearing aid offers controls that enable the hearing aid user to change hearing aid settings in their own listening environments in a parametric way. A memory in the device collects and stores information about user-selected settings together with the acoustic input of the listening environment at the time a selection of hearing aid settings was made. Over time, a training algorithm will relate the preferred hearing aid setting to the acoustic characteristics of the listening environment and begin to make predictions about the hearing aid user's preferred hearing aid settings, depending on the acoustic input (Dillon *et al.*, 2006). It is anticipated that over time, the hearing aid user will train the device to provide individually preferred settings in different listening environments.

The concept is thought to overcome difficulties in fine-tuning a hearing aid to an individual's specific needs in a clinical setting, as the adjustments are made by the hearing aid user in his or her own listening environments (Dillon *et al.*, 2006). With the average age of hearing aid candidates decreasing and the population in general becoming more technology minded, participation in the fitting process is expected to be embraced by hearing aid users. It has also been hypothesized that user involvement in the hearing aid fitting process can result in a greater feeling of ownership of the device and its settings. For example, Schweitzer *et al.* (1999), who suggested moving toward a "consumer-driven psychophysical method of adjustment" that relied less on measurements and more on perceptually driven control, noted that hearing aid users liked to be more directly involved in the fitting process. Similarly, Elberling and Vejby Hansen (1999), who had listeners make adjustments to the gain-frequency response in the laboratory using a three-band equalizer, reported that their participants found it "exciting" to perform the adjustments, and that they appreciated not having to explain to the clinician what they wanted from the amplification. In support of these observations, a recent consumer survey in the United States found that among current nonusers who were considering buying a hearing aid, the ability to self-fit or self-adjust their hearing aids would increase the likelihood of purchase for 19% of those surveyed (Kochkin, 2007). We note that some interactive fitting procedures performed in the clinic with client participation have led to positive outcomes (e.g., Kiessling *et al.*, 1996; Lunner *et al.*, 1997; Moore *et al.*, 1998; Moore *et al.*, 2005).

It is also anticipated that the trained algorithm will deliver more accurate predictions of the preferred hearing aid settings across a wide range of listening environments than will automatic hearing aids that change amplification based on a classification of listening environments according to their acoustic characteristics (Kates, 1995; Nordqvist and Leijon, 2004; Büchler *et al.*, 2005; Hamacher *et al.*, 2005). This is because the acoustic characteristic will vary (e.g., in overall spectral shape) even within a certain class of sounds (e.g., noise), requiring different amplification characteristics to make the environment comfortable to the hearing aid user (Keidser *et al.*, 2005a). Therefore, more precise predictions

[a]Electronic mail: gitte.keidser@nal.gov.au

are expected to result from analyzing listening environments along a continuum of selected acoustic parameters, rather than as discrete classes.

The trainability concept was implemented and evaluated in a body-worn, binaural, programmable research device (Zakis *et al.*, 2004) that provided slow acting wide dynamic range compression and modulation depth based noise suppression in three channels. In this version of a trainable hearing aid, the user accessed one rotary control button that alternated between changing the overall gain (volume), the slope of the gain-frequency response, and the gain at the midfrequencies relative to the gain at the low and high frequencies. When the user had reached a preferred setting, a separate button that logged a "vote" for that setting was pressed. The compression threshold (CT), the gain below the CT, the compression ratio, and the noise suppression strength in each hearing aid channel were trained using information about the intensity level and estimated signal-to-noise ratio overall and in each of the three hearing aid channels, and the preferred gain settings in three channels. Thirteen listeners with a symmetrical, sensorineural hearing loss trained the device starting with a base line response corresponding to that prescribed by NAL-NL1 (Dillon *et al.*, 1999) for one to four weeks. Thereafter, the NAL-NL1 and the trained responses were implemented in two programs of the research device and compared in the field. The comparison trial was conducted double blinded as the two responses were randomly assigned to each of the two hearing aid programs each time the device was switched on. Program preferences were logged using the before mentioned voting button. At the conclusion of the comparison trial, nine listeners showed a significant preference for the trained settings, three listeners showed no significant preference, and one listener significantly preferred the untrained response, which was thought to be because of measurement errors in the fitting process. The study demonstrated that hearing aid users were able to use a complex control configuration to train the hearing aid to produce amplification characteristics that were perceived as better than those resulting from a generic prescription procedure in a variety of listening environments. More details may be found in Zakis *et al.* (2007).

In recent years, a simple form of trainability has been introduced in commercial hearing devices. So far, trainability in commercial products tends to focus on the overall gain based on the hearing aid user's settings of the volume control (Chalupper, 2006; Dijkstra *et al.*, 2007; Phonak, 2006; Oticon, 2007; Unitron, 2007). However, it is likely that more sophisticated training algorithms will find their way into hearing devices in the near future. Although the initial experience reported by Zakis *et al.* (2007) with a more sophisticated trainable hearing aid was positive, a survey of the clinical population has suggested that a trainable device must be simple to manage and that the training must be effectively achieved within a few weeks in order to be of interest to the average consumer (Keidser *et al.*, 2007).

Over the years, work has continued at NAL with the aim of refining the parameters used by Zakis *et al.* (2007). For example, an investigation into the use of different control configurations revealed that participants were able to reliably adjust the gain-frequency response to different preferred settings when presented with various listening environments on audio/video in the laboratory (Dreschler *et al.*, 2008). The fact that the controls manipulated different gain parameters did not affect the end result. However, the listeners did have a preference for two control systems that enabled adjustment of the overall gain and slope of the gain-frequency response, or the overall gain and gain at low and high frequencies. A controller that resembled the one used by Zakis *et al.* (2007) was, in fact, the least preferred and the least reliable of the four. An interesting finding in the study of Dreschler *et al.* (2008), and one that is of concern for any fitting procedure in which client preferences affect the final responses, was a significant effect of the base line (starting) response. In that study, base line responses were tilted away from the NAL-RP prescribed response (Byrne and Dillon, 1986; Byrne *et al.*, 1991) by a randomly selected gain for each octave value, and an additional, random roving of overall gain was applied to the base line. Specifically, the average overall gain and frequency response slopes selected from each base line response were linearly correlated with the starting gain and frequency response slope, respectively. The slopes of the regression lines were close to 2:1. That is, if the base line response had 4 dB less overall gain and a 6 dB/octave steeper slope than prescribed, then listeners would, on average, select 2 dB less gain and a 3 dB/octave steeper slope than selected from the NAL-RP prescribed base line response. It was speculated in that paper that the listeners may have either made conservative adjustments or they may have had a wide range of preferences within a large range of responses. It was also suggested that the effect of the base line response would be of less importance in an actual training paradigm wherein the base line response would move adaptively toward the preferred setting after each adjustment.

This paper describes two experiments. The first experiment investigates the effect of the base line response shape when repeated adjustments are made in a given listening environment and the adjusted setting after each repetition becomes the base line response for the next adjustment. This procedure does not exactly replicate the course of trainability, but provides an insight into the hearing aid user's ability to reach an optimum setting from different starting points when using an adaptive procedure that resembles a training paradigm. To keep the study within a reasonable time frame, the experiment was limited to five adaptive adjustments from two different base line responses using a selection of real-life environments as stimuli. It was also anticipated that five adjustments were sufficient to establish if the participants selected their preferred response immediately or if they continued to move the response in a particular direction. In the second experiment, three responses within the range of preferred responses that resulted from two different base line responses were presented in a paired-comparison test in order to determine if the responses were equally preferred or whether the listeners had a reliable preference for one over the others. The perceptual difference between the responses was also explored. The specific research questions we wished to answer were as follows:

(1) During self-adjustments, will hearing aid users move toward a single optimized response shape when starting from two different adaptively changing base line responses?
(2) In a direct comparison, can hearing aid users reliably discriminate between different response shapes that have been deemed "preferred"?
(3) In a direct comparison, do hearing aid users have a reliable preference for one of the three different response shapes that have been deemed preferred?

II. EXPERIMENT I

A. Methodology

1. Listeners

Twenty-one male and three female listeners participated in the study. The listeners' ages ranged from 18 to 87 years, with a median age of 75 years. All listeners had symmetrical (defined as $\leq$10 dB difference between left- and right-ear pure-tone air conduction thresholds at all audiometric frequencies from 0.25 to 8 kHz, with a difference of 15 dB allowed at one frequency above 1.5 kHz) sensorineural hearing loss. All but one were experienced binaural users of amplification. The length of time with amplification ranged from six months to 15 years. Of the 23 hearing aid users, one wore completely in the canal hearing aids, 10 wore in-the-ear hearing aids, and 12 wore behind-the-ear hearing aids. The pure-tone average (PTA) (average of pure-tone air conduction thresholds at 0.5, 1, and 2 kHz) of the group ranged from 18 to 64 dB HL, with a mean PTA of 40 dB HL. Listeners were selected to represent a variety of hearing loss configurations, with 6 listeners demonstrating a steeply sloping loss [$\geq$40 dB difference between the average of pure-tone air conduction thresholds at 2, 3, and 4 kHz (HFA) and the average of pure-tone air conduction thresholds at 0.25, 0.5, and 1 kHz (LFA)], 13 a gently sloping loss (20–39 dB difference between HFA and LFA), and 5 a flat loss ($<$20 dB difference between HFA and LFA). Participants were paid for their participation in the study. An overview of the listener data is shown in Table I.

2. Stimuli

From a video library of 37 real-life environments recorded by NAL, 12 videos displaying different characteristics across the intensity, spectral, and temporal domains and representing a variety of speech, noise, and speech in noise situations were selected. Table II shows an overview of the stimuli together with 3 of the 23 acoustic parameters used to analyze the intensity, spectral, and temporal characteristics of the sounds. The three parameters include the long-term root-mean-square (rms) level (Leq), the first moment of the frequency spectrum (frequency weighting), and the mark-space ratio (MSR). The first moment of the frequency spectrum is obtained from a 128-point fast Fourier transform spectrum by $\Sigma x_i y_i / \Sigma y_i$, where x is the ith frequency and y_i is the magnitude output in dB for frequency x_i. The outcome is the frequency where the highest intensity levels occur, on average, over time. The MSR is a measure of amplitude variation that equals the time for which the envelope exceeds the mean envelope value divided by the time that the envelope is less than the mean envelope. The more pulsatile the stimulus is, the lower is the MSR value. For real-life sounds, the MSR typically varies between 0 and 1. The stimuli included a soft dialog in a library, a table fan in operation, a female talker reading aloud, hosing down a car, a washing machine in operation, riding in a car with the windows down, a dialog in a café, vacuuming in a living room, applause in a lecture room, a monolog in a car while driving, riding on a train, and using a hair dryer. Across these environments, the real-life presentation levels varied from 50.5 dB SPL (soft dialog) to 71.5 dB SPL (hair dryer) and the first moment of frequency varied from 0.177 kHz (car ride) to 3.257 kHz (hosing down a car). The hair dryer stimulus fluctuated the least in terms of amplitude (MSR=0.81), while applause fluctuated the most (MSR=0.41).

3. Instrumentation

The videos were played back from a Panasonic AG-DV2700 digital video cassette player with the video output directed to an 80 cm Sony television monitor. The stereo audio output was routed through a variable preamplifier (built in-house), two Behringer Ultra-Curve Pro DSP 8024 digital equalizers, a variable and remotely controlled attenuator (built in-house), and a Sony TA-F242 integrated stereo power amplifier to two B&W 600 series loudspeakers. The digital equalizers each enabled gain changes of $\pm$16 dB in 0.5 dB steps overall (master gain) and at each one-third octave frequency from 0.02 to 20 kHz. A frequency response that served to flatten the responses of the equipment used for recording and playback of the stimuli was implemented in one of the equalizers.

One of the digital equalizers was remotely controlled via a personal computer from a keypad that had two rows of three keys. The keys were labeled "volume," "bass," and "treble." The upper row of keys would increase gain and the lower row of keys would decrease gain, in 3 dB steps. Pressing the volume key would first change the master gain level. If +16 or −16 dB was reached, gain was then increased or decreased, respectively, at all frequencies until +16 or −16 dB was reached at one or more individual frequencies. The bass and treble keys changed gains at 0.4 and 4 kHz, respectively. For frequencies below 0.4 kHz and above 4 kHz, the gain change was equal to the gain changes at 0.4 and 4 kHz, respectively. When gain changed at 0.4 or 4 kHz, gain was linearly interpolated in dB on a logarithmic frequency scale between 0.4 and 1.25 kHz or between 1.25 and 4 kHz. If +16 or −16 dB was reached at one or more frequencies between 0.02 and 1.25 kHz, the master gain would increase or decrease, respectively, while at the same time gain would decrease or increase between 1.25 and 20 kHz. If +16 or −16 dB was reached at one or more frequencies between 1.25 and 20 kHz, the master gain would increase or decrease, respectively, while at the same time gain would decrease or increase from 0.02 to 1.25 kHz. This implementation of the controls meant that each key potentially had a 64 dB adjustment range.

A software program developed specifically for the experiment controlled the communication between the keypad

TABLE I. An overview of the listener data.

Listener	Gender	Age	Loss configuration	Ear	HTL 0.25 kHz	HTL 0.5 kHz	HTL 1 kHz	HTL 2 kHz	HTL 4 kHz
306	M	72	Steeply sloping	L	20	30	30	55	75
				R	20	20	25	55	80
309	M	74	Gently sloping	L	35	35	40	60	65
				R	35	30	45	60	70
310	M	74	Gently sloping	L	50	45	45	60	75
				R	60	50	45	65	80
319	M	73	Flat	L	35	40	20	30	65
				R	30	30	25	35	70
328	M	71	Gently sloping	L	30	30	25	40	60
				R	25	20	35	40	60
329	M	76	Gently sloping	L	25	25	25	40	65
				R	20	15	25	40	60
331	M	76	Gently sloping	L	25	45	35	35	75
				R	30	35	30	30	70
333	F	72	Flat	L	35	55	55	55	65
				R	30	55	50	60	60
335	M	84	Gently sloping	L	45	45	35	55	75
				R	50	40	40	55	65
336	M	81	Steeply sloping	L	10	5	10	40	70
				R	15	15	10	25	75
345	M	76	Steeply sloping	L	15	10	30	40	70
				R	0	20	30	45	70
346	M	80	Gently sloping	L	20	15	15	50	70
				R	25	15	20	40	65
347	M	80	Gently sloping	L	35	35	35	60	65
				R	35	35	40	60	65
348	M	61	Steeply sloping	L	20	30	45	65	80
				R	25	35	55	75	95
351	M	87	Gently sloping	L	20	25	30	45	80
				R	35	30	30	45	70
353	M	74	Steeply sloping	L	15	15	40	70	70
				R	10	15	40	65	70
354	M	18	Gently sloping	L	25	25	35	55	55
				R	25	25	35	55	55
356	M	80	Gently sloping	L	25	35	45	50	70
				R	30	35	45	45	55
357	M	79	Gently sloping	L	35	45	75	70	75
				R	40	50	70	75	80
358	M	73	Gently sloping	L	35	30	25	35	70
				R	35	30	30	30	65
359	F	77	Flat	L	35	45	50	60	65
				R	40	45	55	55	60
360	M	83	Steeply sloping	L	30	30	30	60	80
				R	15	25	30	55	70
361	M	65	Flat	L	45	55	65	65	65
				R	40	50	55	65	65
362	F	70	Flat	L	25	30	40	30	45
				R	25	35	30	35	55

and the equalizer. It also accumulated and stored the preferred gain settings from each base line response for each stimulus. The preamplifier and variable attenuators were used to implement the interstimulus differences in Leq. Calibration of the instrumentation was carried out daily.

4. Base line responses

Two base line responses, "flat" and "steep," were implemented for each listener by applying tilts of −4.8 and +4.8 dB/octave, respectively, to the individually prescribed NAL-RP response. Because all hearing losses were symmetrical, the NAL-RP response was prescribed for each individual on the basis of the average hearing loss across the two ears. The tilts were implemented by increasing or decreasing gain by 8 dB at 0.4 and 4 kHz. For frequencies between 0.4 and 4 kHz, the variation was interpolated linearly in dB on a logarithmic frequency scale, resulting in 0 dB gain applied at 1.25 kHz (see Fig. 1). While the

TABLE II. The selected test stimuli and their Leq, frequency weighting (first moment of the frequency spectrum), and MSR.

Stimulus	Leq (dB SPL)	Frequency weighting (kHz)	MSR
1. Soft dialog	50.5	0.87	0.65
2. Table fan	58.5	1.60	0.54
3. Female talker	61.0	1.25	0.54
4. Hosing car	63.5	3.26	0.50
5. Washing machine	64.0	1.15	0.47
6. Car ride	65.0	0.18	0.55
7. Dialog in café noise	72.0	0.86	0.56
8. Vacuuming	72.0	2.07	0.73
9. Applause	73.0	2.08	0.41
10. Monolog in car noise	74.0	0.61	0.66
11. Train ride	77.0	1.47	0.66
12. Hair dryer	77.5	2.35	0.81

NAL-RP component of the base line response was implemented in the digital equalizer that was not under the listener's control, the tilts were implemented in the listener-controlled equalizer to allow for a greater adjustment range in the direction away from the tilting. The rationale for the latter was based on the assumption that the listeners were more likely to undo the changes implemented to the NAL-RP prescription than to move the response further away from the prescription. Further, under an assumption that most of the participants would be wearing nonlinear amplification and therefore at least would change the overall gain for softer and louder stimuli processed with the linear NAL-RP prescription, the Leq values of the test stimuli were revised by reducing or increasing gain on a 2:1 ratio for stimuli with a Leq above or below 65 dB SPL, respectively. The revision to the overall gain was implemented in the non-controlled equalizer.

5. Protocol

Listeners attended the laboratory for three appointments. At the first appointment, listeners were provided with an information sheet that detailed the requirements of the study and were asked to sign a form giving their consent to participate. Otoscopy, tympanometry, and pure-tone air and bone conduction audiometry were conducted for those listeners who had not been tested within the past six months.

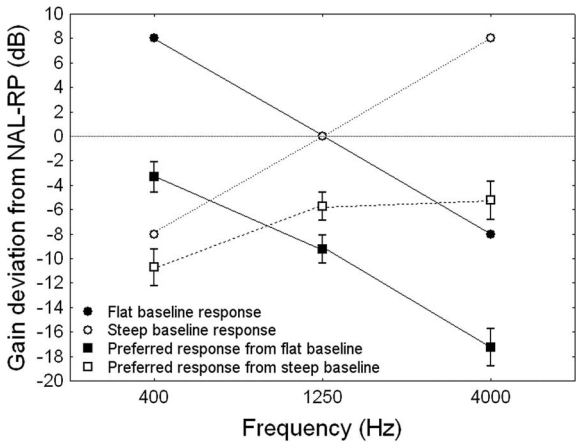


FIG. 1. The gain deviation from NAL-RP of the two base line responses (thin lines) and the average gain deviation from NAL-RP selected at three discrete frequencies after five rounds of adjustments by 24 listeners across the 12 stimuli when starting from the two base line responses (thick lines). The bars show the 95% confidence bands.

The self-adjustment task took place in a 10×6 m^2 quiet room with a low reverberation time that simulates a domestic living room. During testing, listeners were seated facing the television monitor at a distance of 1.8 m, with the two loudspeakers facing the listener from angles of approximately $\pm 20°$ on opposite sides of the monitor. An acoustically transparent curtain covered the loudspeakers. Stimuli were presented in the free field to the unaided listeners. First, a loudness scaling task was completed to ensure that listeners did not encounter uncomfortable loudness levels when the stimuli were presented using the initial base line responses. Each listener was presented with the hair dryer stimulus in the free field at a level that was 20 dB lower than the intended presentation level using his/her individual prescription and was instructed to use a seven-point categorical scale to rate its perceived loudness (Cox *et al.*, 1997). The level of the stimulus was increased in 5 dB steps until the listener rated the sound as "loud, but ok," after which the level was increased in 2 dB steps until the listener rated the sound as "uncomfortably loud" or the intended presentation level was reached. Three test runs were completed for each base line response. If a listener found the stimulus uncomfortably loud at or below the intended presentation level, the levels at which the listener first rated the sound as *loud, but ok* during the third run for each base line response were averaged, and subsequent stimuli were presented at or below this level. The initial presentation level of the highest level stimuli was reduced for all listeners, and the presentation level was reduced for all stimuli for two listeners. The initial gain reduction did not, however, prevent the listeners from choosing higher gain levels during the experiment. Instructions for the loudness scaling task are shown in Appendix A.

Listeners were then presented with each of the 12 video stimuli and asked to use the keypad to adjust the amplification characteristics from one of the two base line responses to their preferred setting. Instructions for the adjustment task are shown in Appendix B. Two adjustments for each base line response were completed. For each stimulus, the average of the two adjusted settings for each base line response served as the base line responses for the second appointment (round 2).

At the second appointment, listeners again used the keypad to adjust the amplification characteristics to their preferred setting; this time using the average adjusted responses from round 1 as the base line responses. Two adjustments (rounds 2 and 3) from each base line response were completed. The adjusted settings from round 2 served as the round 3 base line responses, from which the listeners completed another set of adjustments.

At the third appointment, listeners repeated the procedure from the second appointment (rounds 4 and 5), using the adjusted settings from round 3 of the second appointment as the first set of base line responses. The adjusted settings from round 4 served as the round 5 base line responses.

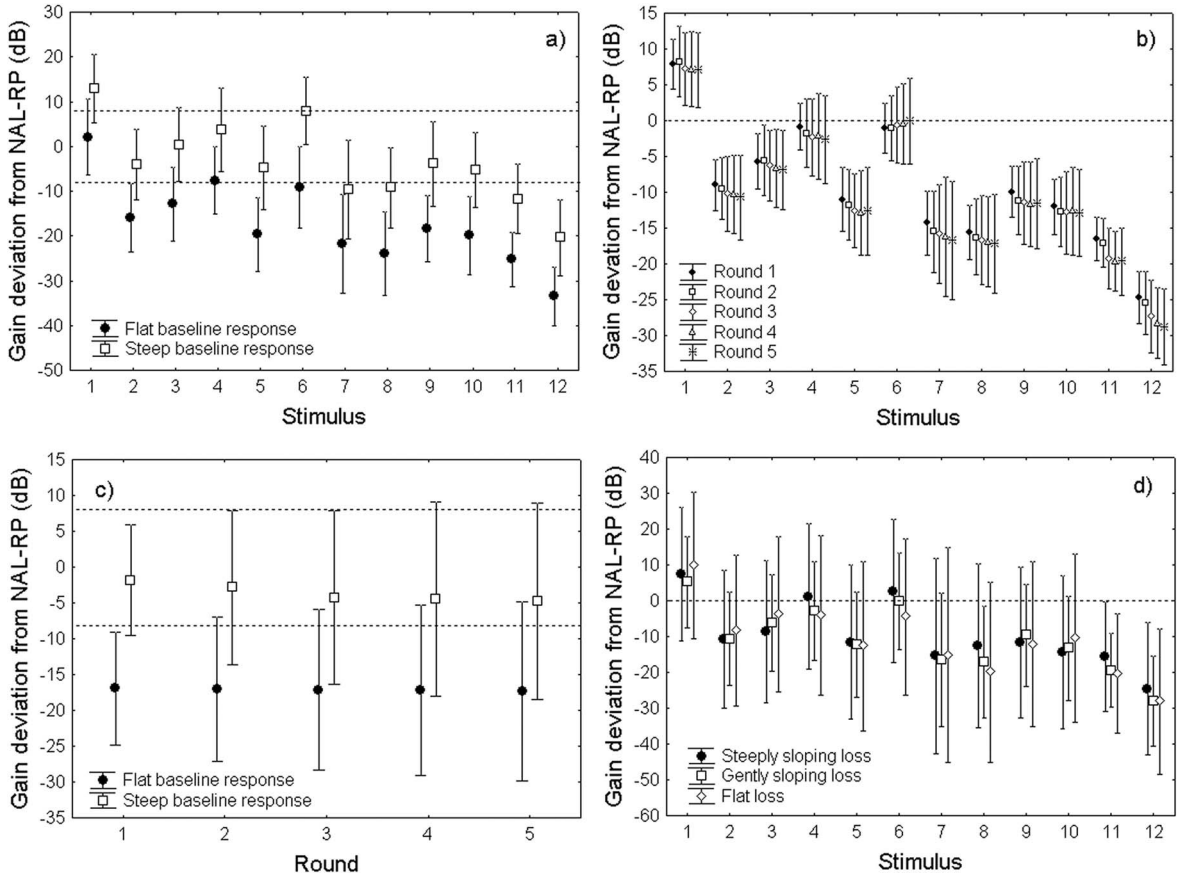


FIG. 2. The average gain deviation from NAL-RP selected at 4 kHz shown in interactions between (a) stimulus and base line response, (b) stimulus and round, (c) round and base line response, and (d) stimulus and hearing loss configuration. The stimulus numbers in graphs a, b, and d correspond to the numbers in Table II. The broken lines in graphs a and c show the starting level at 4 kHz from the flat (−8 dB) and steep (+8 dB) base line responses.

The order of stimulus presentation and the initial base-line response was determined by a balanced Latin square.

B. Results

From the gain settings reached after each round of adjustments from each of two base line responses, the selected gain relative to the NAL-RP prescription at 0.4, 1.25, and 4 kHz was derived for each listener and stimulus.

1. Group performance

A repeated measures analysis of variance on the selected gain deviation from NAL-RP using stimulus, base line, and round as repeated measures and hearing loss as the between-group measure was conducted separately for each frequency (0.4, 1.25, and 4 kHz). Of the main effects, stimulus and base line were highly significant at all three frequencies ($F_{11,231}>85.2$, $p<0.000\,000\,1$ and $F_{1,21}>87.4$, $p<0.000\,000\,1$, respectively). According to Tukey HSD *post hoc* analyses, significantly different gain deviations from the prescription were selected across many stimuli. At each frequency, about 30 dB less gain was preferred for listening to the hair dryer (stimulus 12) than the soft dialog (stimulus 1), see Fig. 2(b). With respect to the base line response effect, the gain selected from the flat base line response varied from the gain selected from the steep base line response by, on average, 8.7, −3.1, and −12.9 dB at 0.4, 1.25, and 4 kHz, respectively. Figure 1 shows the average selected gain deviations from NAL-RP for the flat and steep base line responses across all stimuli after the last round of adjustments. It can be seen that the flat response shape was very much maintained in the average selected response after the final adjustment, while relatively less gain was selected at the high than at the low and midfrequencies when starting from the steep base line response. In addition, round reached significance for 4 kHz ($F_{4,84}=2.7$, $p=0.04$), at which frequency the listeners selected significantly lower gain during rounds 3, 4, and 5 (<−10.9 dB with regard to NAL-RP) than during round 1 (−9.4 dB with regard to NAL-RP), see Fig. 2(c).

At all three frequencies, there were significant two-way interactions between stimulus and base line ($F_{11,231}>161$, $p<0.0002$), stimulus and round ($F_{44,924}>264$, $p<0.003$), base line and round ($F_{4,84}>12$, $p<0.02$), and stimulus and hearing loss ($F_{22,231}>1.6$, $p<0.04$). The interactions at 4 kHz are illustrated in Fig. 2. According to Tukey HSD *post hoc* analyses, the base line effect was significant for all stimuli at 0.4 and 4 kHz at which frequencies gain could be independently adjusted [see Fig. 2(a)]. At 1.25 kHz, the base line effect was significant for all but four stimuli (soft dialog, female talker, hosing car, and hair dryer) for which the average preferred gain from each base line response was within 2.2 dB. The effect of round was significant for a handful of stimuli representing mainly noise-only situations such as

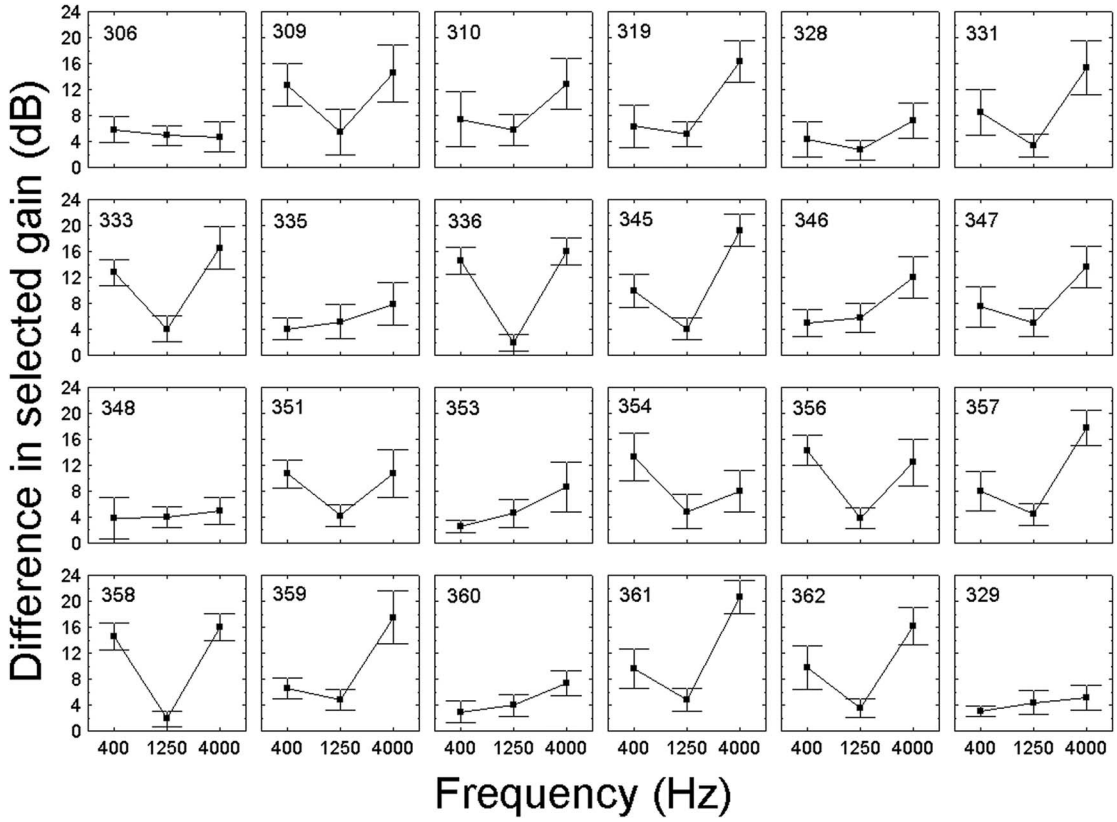


FIG. 3. The average absolute difference in selected gain measured at three discrete frequencies from two base line responses after five adjustment rounds by each listener. The bars show the 95% confidence bands.

table fan, washing machine, train ride, and hair dryer [Fig. 2(b)]. At 0.4 kHz, the participants, on average, tended to reduce gain over rounds when starting from the flat base line response and to increase gain over rounds when starting from the steep base line response. At 1.25 kHz, gain was, on average, increased when starting from the flat base line response, while at 4 kHz, gain, on average, was reduced when starting from the steep base line response. In all cases, the gain deviation selected at round 1 was significantly different from the gain deviations selected from round 3 onward [see Fig. 2(c)]. Because the average gain adjustment made during the first round at all three frequencies (>4.3 dB) clearly exceeds the gain adjustments made between successive rounds (mostly <1 dB), the hypothesis raised in Dreschler *et al.* (2008) that the base line response biases seen after one adjustment are due to the conservatism of listeners in their initial adjustments may be rejected. Finally, at none of the frequencies did the *post hoc* analysis reveal a significant difference in the gain deviations from NAL-RP selected for each stimulus by listeners with different hearing loss configuration [see Fig. 2(d)]. As the NAL-RP prescription for individual listeners takes the hearing loss at each frequency and the slope of the hearing loss into account, this finding suggests that the preferred gain deviations are determined by changes in the acoustic characteristics of the stimuli, rather than the individual's audiometric configuration.

There was further a significant three-way interaction between hearing loss, stimulus, and round at 1.25 kHz ($F_{88,924}=1.3$, $p<0.03$), and between hearing loss, base line, and round at 4 kHz ($F_{8,84}=43.5$, $p<0.006$). No further interactions showed significance. Tukey HSD *post hoc* analyses revealed that at 4 kHz, participants with sloping (steeply or gently) hearing loss selected significantly lower gain from the steeper base line response during rounds 4 and 5 than during the first two rounds. The effect of round at 1.25 kHz was significant for more stimuli when the hearing loss was steeply than gently sloping. There was no significant effect of round for participants with flat hearing loss. These results suggest that participants with sloping hearing loss were more likely to keep fine-tuning the preferred response than were participants with flat hearing loss.

2. Individual results

Figure 3 shows for each participant the average difference between the final gains selected at 0.4, 1.25, and 4 kHz from the two base line responses across the 12 stimuli. From the gain difference values at these three frequencies, the rms difference between the preferred responses ($\Delta\text{rms}_{\text{pref}}$) was calculated. Across the participants the average $\Delta\text{rms}_{\text{pref}}$ varied from 4.7 to 13.7 dB, with a mean of 9.2 dB. For each participant, the slope from 4 to 0.4 kHz of the gain difference curve ($\Delta\text{Slope}_{\text{pref}}$) was also obtained. The average $\Delta\text{Slope}_{\text{pref}}$ varied from −1.8 to 3.3 dB/octave, with a mean of 1.2 dB/octave. Both $\Delta\text{rms}_{\text{pref}}$ and the absolute value of $\Delta\text{Slope}_{\text{pref}}$ were highly and significantly correlated with the average absolute difference in gain selected at 4 kHz ($r=0.82$, $p<0.0001$ and $r=0.62$, $p=0.001$ for $\Delta\text{rms}_{\text{pref}}$ and $\Delta\text{Slope}_{\text{pref}}$, respectively), suggesting that the variation in the preferred responses was largely explained by how well the

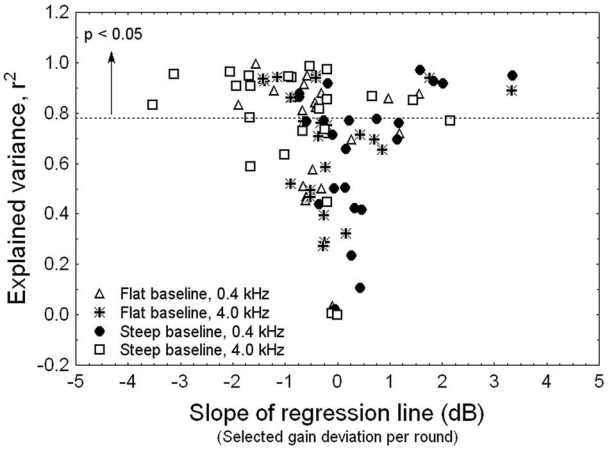


FIG. 4. The relationship between the variance explained by the regression line fitted to the average gain adjustment per round (after round 1) produced by each study participant as a function of round, and the slope of the regression line. The slope of the regression line indicates how much the participant continued to change gain from round 1 to round 5, and the explained variance is a measure of the randomness of the observed gain changes. A higher value shows less randomness.

listeners were able to choose their preferred gain for the high frequencies. There was a further moderate, but significant, correlation between the average absolute difference in gain selected at 4 kHz from the two base line responses and the slope of hearing loss ($r=-0.49$, $p=0.02$) as listeners with more sloping high-frequency hearing loss selected more consistent gain at 4 kHz from the two different base line responses (e.g., listeners 306, 348, and 360 in Fig. 3). This observation suggests that listeners with sloping high-frequency hearing loss were more certain than were listeners with flat hearing loss about their preferred high-frequency gain.

The gain adjustment with successive rounds also varied among the participants. For each participant, a regression line was fitted to the average gain selected across stimuli per round (after round one) as a function of round. This was done for data obtained for each of the frequencies 0.4 and 4 kHz, for which gain could be independently adjusted, and for each base line response. Figure 4 shows for each listener the relationship between the proportions of variance accounted for in the gain changes by round and the slope of the regression line. For most participants, the average gain adjustment from one round to the next was less than 1 dB (slope between +1 dB and −1 dB). For some of these participants, the tendency to keep changing gain at this rate was strong and statistically significant (at and above the dashed line), while for others, the adjustments by round were random (below the dashed line). A small proportion of listeners (primarily with steeply sloping hearing loss) continued to change gain by more than 1 dB (up to 3.5 dB) per round in a systematic way, more so at 4 kHz and when starting from the steep base line response.

To explore further the variation in individual performance, a second experiment was designed in which a subset of listeners from experiment I compared the average preferred response shape from each base line response with one that had the shape toward which the group of listeners appeared to be moving through the adaptive adjustment procedure.

III. EXPERIMENT II

A. Methodology

1. Listeners

Ten male and two female listeners from experiment I (309, 319, 328, 329, 331, 333, 336, 345, 348, 351, 359, and 360) participated in a paired-comparison test. The subset of 12 listeners was selected to represent a variety of response ranges within which the selected responses from two different base lines fell in the previous experiment. Three listeners (329, 348, and 360) produced a narrow response range ($\Delta rms_{pref}<6$ dB), five listeners (319, 328, 331, 351, and 359) produced a medium response range (Δrms_{pref} between 6 and 12 dB), and the remaining four listeners produced a wide response range ($\Delta rms_{pref}>12$ dB). A variety of hearing loss degrees and audiometric configurations was also sought. All listeners in the group had symmetrical sensorineural hearing loss, including five listeners with gently sloping hearing loss, four listeners with steeply sloping hearing loss, and three listeners with flat hearing loss. The PTA ranged from 18 to 55 dB HL, with a mean of 37 dB HL.

2. Instrumentation

The instrumentation was the same as in experiment I but with a third digital equalizer included to enable implementation of all the desired response shapes. The keypad was configured with a single, horizontal row of three buttons that were labeled "A," "B," and "SEL." A press of the A or B button by the listeners brought up a predetermined response shape on the computer-controlled digital equalizer, while the SEL button was used to indicate a preference for either the A or B response.

3. Stimuli

Six stimuli from experiment I were used: car ride, train ride, female talker, soft dialog, monolog in car noise, and dialog in café noise. Each presentation of the stimuli was filtered through one of three different response shapes, all of which were based on the results of experiment I. The flat response shape had the average (across listeners) preferred gain variation from the NAL-RP prescription that was chosen when the base line response was flatter than NAL-RP. Similarly, the steep response shape had the average selected gain variation from the NAL-RP prescription that was chosen when the base line response was steeper than NAL-RP. Finally, a response shape named approach applied the average variation from the NAL-RP prescription that the listeners, on average, seemed to be approaching from both base line responses. The gain variation at each frequency for the approach response shape was determined by the intersection of two lines fitted to the average gain changes observed from round 1 to round 5 in experiment I. The resulting gain variations were thus not necessarily found between the gain varia-

TABLE III. The absolute rms difference in dB between compared responses for each stimulus.

Stimulus	Flat-steep	Flat-approach	Steep-approach
Soft dialog	6.8	1.3	5.9
Female talker	7.9	2.1	7.1
Car ride	9.2	5.3	3.9
Dialog in café noise	6.6	1.9	5.2
Monolog in car noise	7.7	1.9	5.9
Train ride	8.7	3.8	10.2

tions applied for the flat and steep responses. The gain variations of flat, steep, and approach varied across stimuli but did not vary across listeners, although the actual response shapes that included the individual NAL-RP prescription did vary across listeners. The rms differences between the responses compared (Δrms_{comp}), calculated from the gain differences at 0.4, 1.25, and 4 kHz, are listed in Table III. Overall, Δrms_{comp} between the flat and steep response shapes (7.8 dB on average) was greater than Δrms_{comp} between the flat and approach (2.7 dB on average) and between the steep and approach (6.4 dB on average) response shapes.

4. Protocol

Each listener attended the laboratory twice. At each session, the listeners were asked to compare the three response shapes for each stimulus in a paired-comparison procedure. Each pair (flat-steep, flat-approach, and steep-approach) was presented ten times, for a total of 180 comparisons over both sessions, with stimuli and pairs presented in a randomly determined order. Each response shape was randomly assigned to the A and B buttons on the keypad. Listeners were instructed to listen to each setting and to select the one they would prefer if the same listening situation were encountered in real life (forced choice). A selection was made by leaving the stimulus playing on the preferred response shape (A or B) and pressing the SEL button. After making their selection, listeners indicated the perceived degree of difference between the two settings by checking one of four choices on a questionnaire: "greatly different," "moderately different," "somewhat different," or "not different."

B. Results

For each listener, the number of times each response shape was preferred over each other was recorded for each stimulus. According to the binomial distribution, there is a significant preference for one response shape over another if it is selected at least nine times out of ten presentations. The four questionnaire categories were assigned values ranging from 0 (not different) to 3 (greatly different) and the average degree of difference was calculated for each listener, stimulus, and response pair.

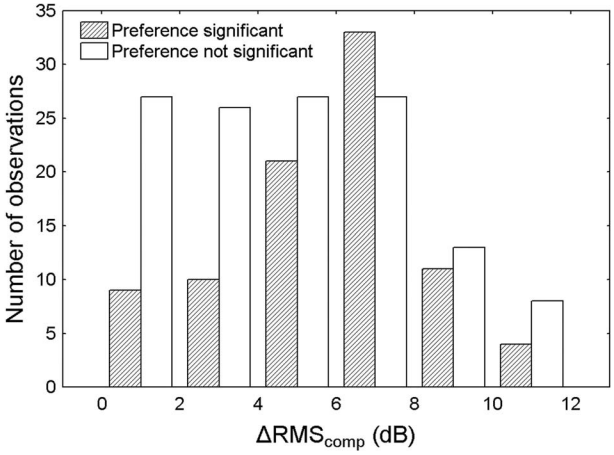


FIG. 5. The distribution of paired comparisons for which the preference for one response shape over the other was significant (shaded bars) or nonsignificant (clear bars) across the rms difference between the responses compared.

1. Preference

Table IV lists the percentage of cases where each of the three response shapes (flat, steep, and approach) was significantly preferred to each of the others for each stimulus and overall (last row). The last column shows the percentage of comparisons in which a significant preference was not reached. For five out of six stimuli, the flat response shape was clearly more often preferred than the steep and the approach response shapes, with the steep response shape being preferred the least number of times. There were more significant preferences for the steep than for the approach response shape for only the dialog in café noise, with only a small number of significant preferences for the flat over the steep response shape. On average, no significant preference was shown in 59.2% of comparisons, so there was not always a clear preference for one response shape over the others.

Grouping the Δrms_{comp} values in 2 dB steps, Fig. 5 shows, for each Δrms_{comp} interval, the proportion of paired comparisons for which listeners significantly chose one response over the other and the proportion of paired compari-

TABLE IV. Percentage of cases in which a response was significantly preferred to the others for each stimulus, response pair, and overall.

Stimulus	Flat-approach		Flat-steep	Steep-approach		No preference
Soft dialog	5.6		16.7	2.8	13.9	61.1
Female talker	8.3		27.8		22.2	41.7
Car ride	5.6	2.8	13.9		11.1	66.7
Dialog in café noise	2.8		5.6	11.1	8.3	72.3
Monolog in car noise	16.7		19.4	5.6	8.3	50.0
Train ride	5.6	2.8	16.7	2.8	8.3	63.9
Overall	**7.4**	**0.9**	**16.7**	**3.7**	**12.0**	**59.2**

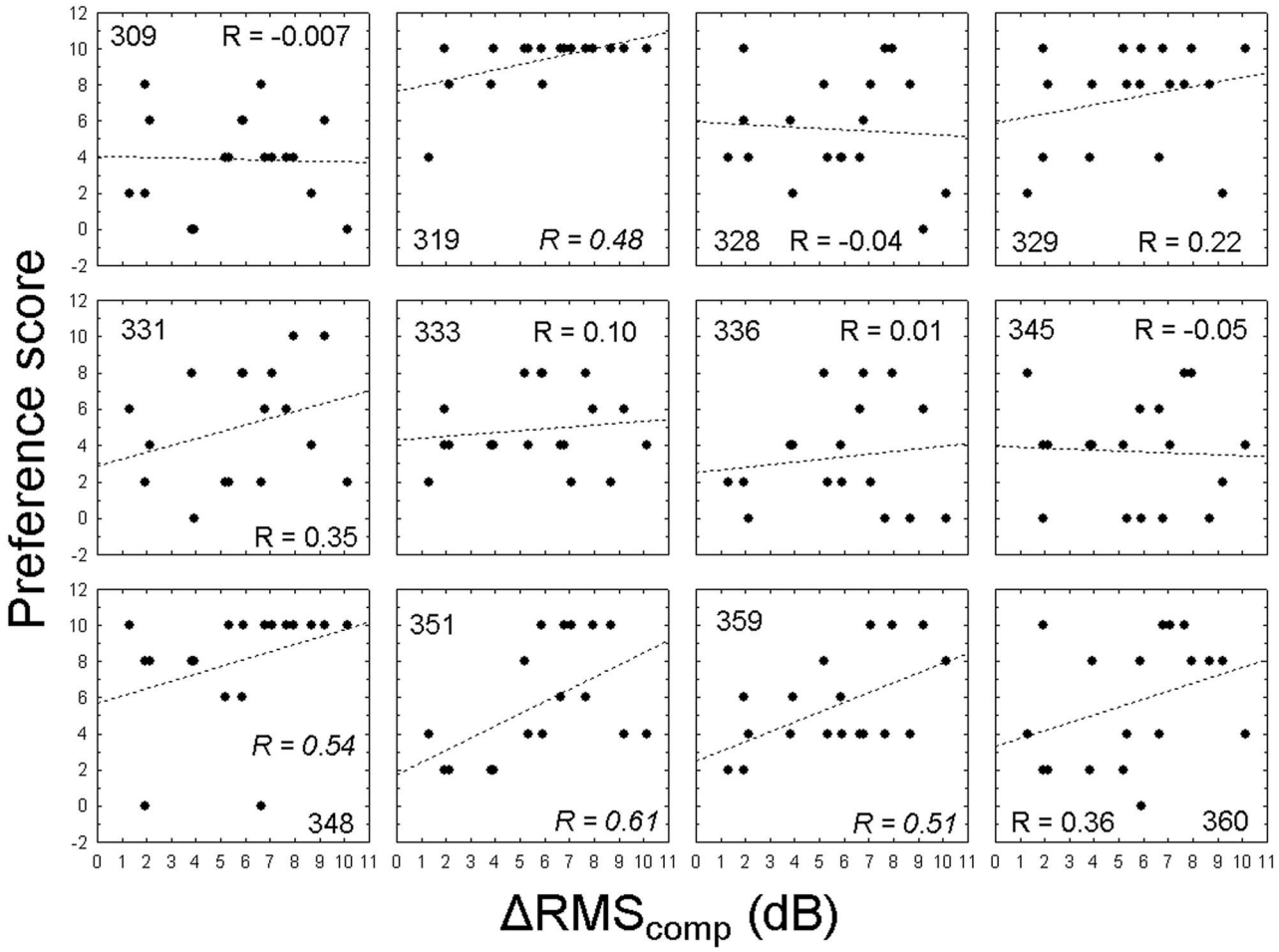


FIG. 6. The relationship between the absolute difference in preference score between two response shapes and the rms difference between the two response shapes for each listener. The broken line shows the regression line, and Spearman R correlation coefficients shown in italics were significant at the 5% level.

sons for which the listeners did not show a significant preference. The distribution of paired comparisons suggests that, even when Δrms_{comp} was small (<4 dB), some listeners could reliably select one response shape over the others. Likewise, even when Δrms_{comp} was large (>8 dB), some listeners could not reliably select one response shape over the other. Thus, there appears to be individual variation in the ability to reliably choose one response over the other, irrespective of the rms difference between response shapes.

A correlation analysis was performed independently for each listener on the preference score (defined as the absolute difference between the number of preferences allocated to two response shapes) and Δrms_{comp}. For example, if there were three preferences for one response shape and seven preferences for the other, the absolute difference in the number of preferences was 4. The result is shown in Fig. 6. Preference scores of 10 or 8 represent significant preferences. It is evident from the mostly flat regression lines that, for most listeners, Δrms_{comp} had little or no influence on the ability to give a clear preference for one response over the other. In particular, three listeners (319, 329, and 348) usually chose their preferred responses reliably irrespective of Δrms_{comp}. Notably, two of these three listeners (329 and 348, both of whom had sloping hearing loss) selected final responses from the two base line responses in experiment I that, on average, were close to each other (Δrms_{pref}=4.7 and 4.9 dB, respectively). A Spearman R correlation analysis showed a high and significant correlation between the average preference score and Δrms_{pref} produced by each participant ($R=-0.81$, $p<0.05$). The smaller the value of Δrms_{pref}, the more likely the listener was to have a significant preference for one of the response shapes compared (Fig. 7). The ability to reliably choose between responses could not be predicted from other parameters, such as age or the audiometric data ($R<0.23$, $p>0.05$). The implication seems to be that only the listeners who were least biased by the base line responses reliably showed a preference for one response shape over the others.

2. Perception

Figure 8 shows for each listener the relationship between their discrimination scores and Δrms_{comp}. Correlation analyses on the data in Fig. 8 revealed that most listeners demonstrated a significant tendency to perceive responses as increasingly different from each other as Δrms_{comp} increased, even though many of these listeners had no clear preference for a single response within a wide range of response shapes (Fig. 6). The only two listeners who did not follow this trend were listener 333, who did not show any reliable relationship

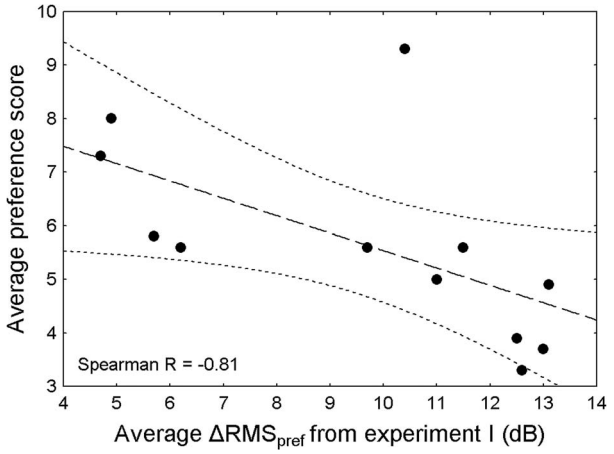


FIG. 7. The relationship between the average preference score and the rms difference between the selected response shapes from two different base line responses in experiment I. The long-dashed line shows the regression line and the short-dashed lines show the 95% confidence bands.

between the two parameters, and listener 336, who rated nearly all responses as not different. Both of these listeners chose final responses from two base lines in experiment I that were far apart (Δrms_{pref}=13.1 and 12.6 dB, respectively).

IV. DISCUSSION

A previous experiment in which listeners made one adjustment from different base line responses demonstrated a strong effect of base line response shape on the preferred response shape (Dreschler *et al.*, 2008). Experiment I in the current study demonstrated that, on average, the effect of base line response was still significant after five adaptive adjustments of gain during which the base line response changed from one adjustment to the next based on the previously selected response shape. However, for some stimuli, listeners tended to continue to adjust the response shape from two different base line responses toward a point between the extremes, more so for individuals whose final responses after five rounds of adjustments were chosen within a narrower range of responses. Thus, despite insignificant gain changes after the third round of adjustments (which suggests that the gain changes had reached a plateau), it is currently unknown whether further adjustments beyond round five would have continued to reduce Δrms_{pref} for some or all listeners.

The ability to select an optimum gain setting from two different base line responses clearly varied across listeners, especially at the high frequencies. Generally, data from this study and a past study (unpublished data) suggest that there was a tendency for listeners to be less biased by a "steeper"

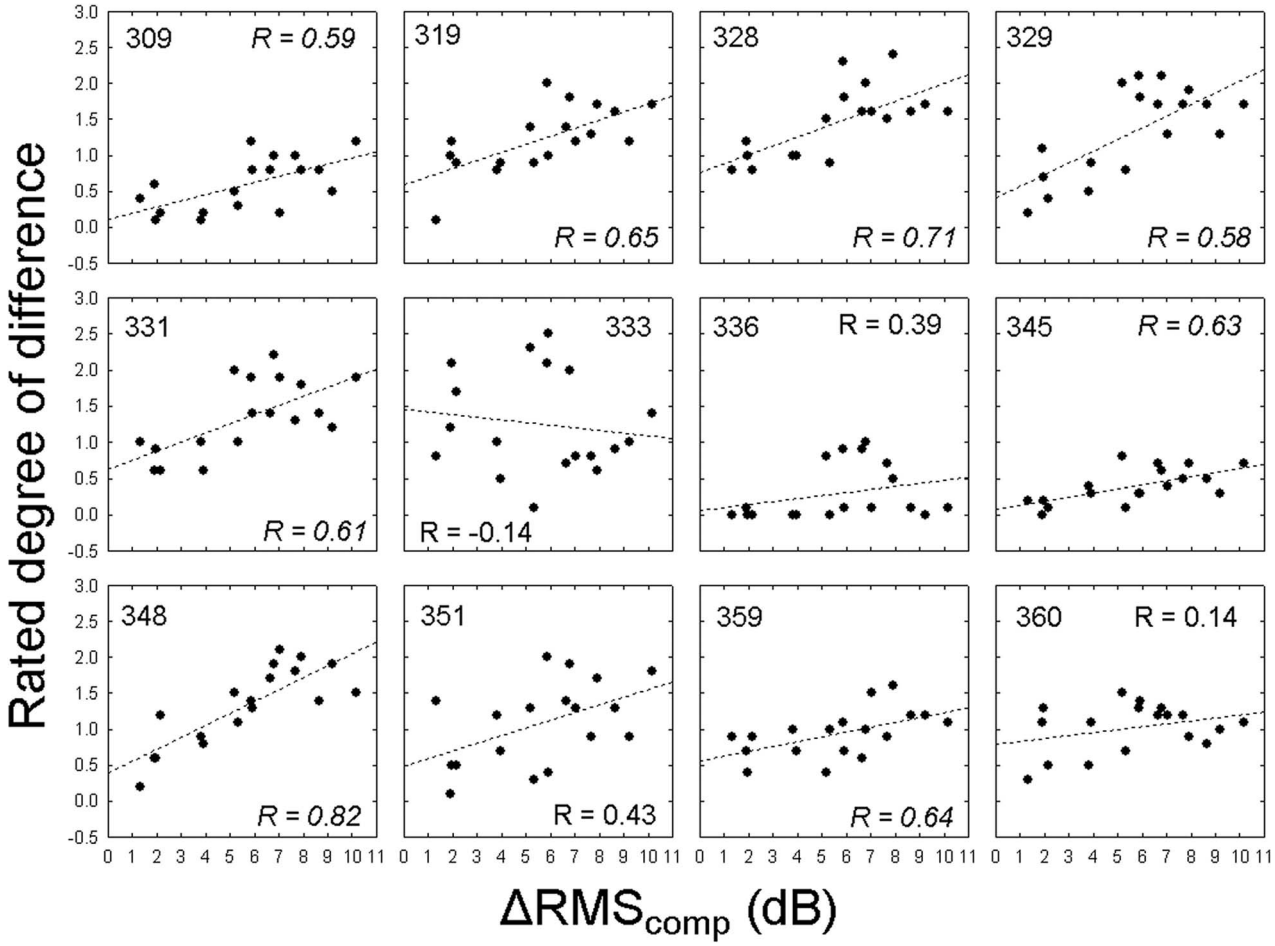


FIG. 8. The relationship between the rated degree of difference between two response shapes and the rms difference between the two response shapes for each listener. The broken line shows the regression line, and Spearman R correlation coefficients shown in italics were significant at the 5% level.

response shape than by a "flatter" response shape relative to NAL-RP. Specifically, the listeners reduced the gain more across the high than across the low frequencies when starting from a steeper base line response. The paired comparison test in experiment II also showed that, when a listener had a significant preference for one biased response shape over the other, the preference was more often for the response shape arrived at from the flatter base line response. This suggests that it would be undesirable to allow self-adjustments from totally randomly selected base line responses, especially as the participants in this study were drawn from a selected list of research volunteers that were likely to better manage the self-adjustment task than the average 75 year old hearing center client. The results also suggest that the listeners were less willing to accept high-frequency gain above than below the NAL-RP prescription. The latter observation is consistent with previous findings that many hearing aid users prefer less high-frequency gain than prescribed by NAL-RP (Keidser, 1995; Keidser *et al.*, 2005a; Magnusson *et al.*, 2001) and NAL-NL1 (Keidser *et al.*, 2005b; Zakis *et al.*, 2007).

Listeners with steeply sloping high-frequency hearing loss were particularly able to narrow the range of preferred responses from two base line responses. The mean Δrms_{pref} for this group was 8.1 dB, while it was 9.3 dB for listeners with gently sloping hearing loss and 12.0 dB for listeners with flat hearing loss. A higher reliability in selecting preferred amplification characteristics among hearing aid users with steeply sloping loss has also been observed in past studies at NAL (Keidser, 1995; Keidser *et al.*, 2005a). The more specific preference shown by this listener group has been explained by their narrower dynamic range at high frequencies, meaning that there is a very clear perceptual effect when the high-frequency gain is adjusted, presumably as sound is moved in and out of their audible range. The between-participant standard error of the mean Δrms_{pref} was larger for listeners with steeply sloping hearing loss than for listeners with gently sloping and flat hearing loss (1.7 dB vs 0.7 and 0.6 dB, respectively), suggesting that within the group of listeners with steeply sloping hearing loss, some were much better than others in narrowing the range of preferred responses from two base line responses. This variation could be related to the individual's ability to extract and utilize information presented at high frequencies (Hogan and Turner, 1998; Ching *et al.*, 1998). For example, some listeners with severe high-frequency hearing loss do not extract useful information from the high frequencies due to a cochlear dead region (Moore, 2001; Moore, 2004).

The ability to discriminate between different responses may also be affected by individuals' cognitive processing skills. Lunner (2003) found that among 17 hearing aid users who had either a high or a low working-memory capacity, those identified with a high capacity were better able to identify and discriminate between different parameter settings in an experimental hearing aid. Out of the 12 listeners who participated in experiment II in the current study, there were three listeners for whom the discrimination between responses in terms of both preference and perception was quite reliable, even when Δrms_{comp} was small (<4 dB). Two listeners, however, did not seem to be able to discriminate between responses either preferentially or perceptually, even when Δrms_{comp} was large (>8 dB). The remaining seven listeners could not reliably choose between responses, but were able to perceive an increasing difference between responses as the rms difference between them increased. Not surprisingly, listeners who demonstrated an ability to consistently choose responses were less affected by the initial base line response shape. Though both tasks rely on auditory processing ability, it may be that producing consistent preferences demands more of an individual's cognitive processing ability (e.g., Pichora-Fuller and Sing, 2006).

We have speculated that the poor preference consistency among many listeners with wide dynamic ranges may have been affected by how well the listeners could relate to the listening situations presented in the laboratory environment. There are examples in the literature demonstrating that presentation of recorded real-life situations in the laboratory produces results different from those obtained in similar situations in the field. For example, Filion and Margolis (1992) found that listening environments presented in the laboratory and in the field at the same sound pressure level were perceived by the listeners as louder in the laboratory. Punch *et al.* (1994) demonstrated that the amplification characteristics preferred under laboratory conditions were only fair predictors of preferences in real life, and Kirkwood (2005) demonstrated that larger errors were made in estimating the lengths of rods dropped onto a floor from auditory input alone when presented through a binaural recording and playback system than when presented live. The lack of realism in the stimuli presented in our experiments may mean that the listeners accepted a wider range of responses than they would in real life, where the surrounding environment often is dynamic and introduces many distracters. From the data presented in this paper, however, it is not possible for us to reject the hypothesis that there is no single optimum response shape within a wide range of responses.

Overall, observations in our laboratory suggest that a self-adjusted response shape depends strongly on the initial response shape, especially at low frequencies and especially for listeners with flat and gently sloping hearing loss. Possible reasons for this finding are that (a) the test environment did not provoke the listeners to seek the optimum response, (b) the shape of the response was not critical within a large range, and/or (c) the listeners were heavily influenced by the starting point, but have a narrower range of preferred response shapes that will take more than five adjustments to reach through self-adjustments. Future studies of self-adjustment of hearing aid gain and trainability should focus on real-life performances and possibly include measurements of participants' cognitive processing abilities.

V. CONCLUSION

There is considerable variation in the ability of hearing-impaired people to select a preferred response shape through adaptive self-adjustments of gain when starting from two different base line responses in the laboratory. Listeners with steeply sloping high-frequency hearing loss were less biased by the initial response shape, especially at high frequencies.

Despite being able to reliably identify perceptual differences between the biased selected responses, only 25% of listeners could demonstrate reliable preferences for one biased response over another. Consequently, choosing an appropriate starting point for self-adjustment of hearing aid gain seems important. The results obtained in this study could be affected by a lack of realism and demand of the laboratory listening environment, but this lack of realism would also apply to any adjustment of hearing aids in clinical situations. It is still uncertain to what extent self-adjustments of hearing aid gain-frequency responses in real-life situations, as required in trainable hearing aids, are dependent on the response initially prescribed into the hearing aids.

ACKNOWLEDGMENTS

This study was supported by the Cooperative Research Center for Cochlear Implant and Hearing Aid Innovation. The authors thank Scott Brewer for the software development, and Brian Moore and an anonymous reviewer for helpful comments on a previous version of this manuscript.

APPENDIX A: INSTRUCTIONS FOR LOUDNESS SCALING

Later in this experiment, you will listen to a number of sounds that vary in loudness. As I do not want you to experience sounds that are uncomfortably loud, this test is to determine the level at which sounds become uncomfortably loud to you.

When you hear the sound, use the scale in front of you to decide which loudness category it falls into. I will gradually increase the volume until you rate the sound as *uncomfortably loud* (category 7), then the sound will stop. We will repeat this procedure several times.

You will hear the sound again later in the experiment. However, it will be presented at a level that is below your uncomfortable loudness level.

APPENDIX B: INSTRUCTIONS FOR STUDY TASK

You will see and listen to a number of video recordings of common listening situations. Your task is to use the buttons on the keypad in front of you to adjust the sound until your preferred setting is reached.

Each videotape will run continuously until you have made your choice. Spend some time experimenting with the buttons and listen to the changes it can make to the sound. There are no "right" or "wrong" settings, and you can take as long as you need to make the adjustments.

Büchler, M., Allegro, S., Launer, S., and Dillier, N. (**2005**). "Sound classification in hearing aids inspired by auditory scene analysis," EURASIP J. Appl. Signal Process. **2005**, 2991–3002.

Byrne, D., and Dillon, H. (**1986**). "The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid," Ear Hear. **7**, 257–265.

Byrne, D., Parkinson, A., and Newall, P. (**1991**). "Modified hearing aid selection procedures for severe/profound hearing losses," in: *The Vanderbilt Hearing Aid Report II*, edited by G. A. Studebaker, F. H. Bess, and L. Beck (York, Parkton, MD), pp. 295–300.

Chalupper, J. (**2006**). "Changing how gain is selected: The benefits of combining datalogging and a learning VC," Hear. Res. **13**, 46–55.

Ching, T. Y. C., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Cox, R. M., Alexander, G. C., Taylor, I. M., and Gray, G. A. (**1997**). "The contour test of loudness perception," Ear Hear. **18**, 388–400.

Dillon, H. (**1999**). "NAL-NL1: A new prescriptive fitting procedure for nonlinear hearing aids," Hear. J. **52**, 10–16.

Dillon, H., Zakis, J., McDermott, H., Keidser, G., Dreschler, W., and Convery, E. (**2006**). "The trainable hearing aid: What will it do for clients and clinicians?" Hear. J. **59**, 30–36.

Dijkstra, T. M. H., Ypma, A., de Vries, B., and Leenen, J. R. G. M. (**2007**). "The learning hearing aid: Common-sense reasoning in hearing aid circuits," Hear. Res. **14**, 34–53.

Dreschler, W. A., Keidser, G., Convery, E., and Dillon, H. (**2008**). "Client-based adjustments of hearing-aid gain: the effect of different control configurations," Ear Hear. **29**, 214–227.

Elberling, C., and Vejby Hansen, K. (**1999**). "Hearing instruments: Interaction with user preference," in *Auditory Models and Non-Linear Hearing Instruments, Proceedings of the 18th Danavox Symposium*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen, pp. 341–357.

Filion, P. R., and Margolis, R. H. (**1992**). "Comparison of clinical and real-life judgments of loudness discomfort," J. Am. Acad. Audiol **3**, 193–199.

Hamacher, V., Chalupper, J., Eggers, J., Fischer, E., Kornagel, U., Puder, H., and Rass, U. (**2005**). "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," EURASIP J. Appl. Signal Process. **18**, 2915–2929.

Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Kates, J. M. (**1995**). "Classification of background noises for hearing-aid applications," J. Acoust. Soc. Am. **97**, 461–470.

Keidser, G., Convery, E., and Dillon, H. (**2007**). "Potential users and perception of a self-adjustable and trainable hearing aid: a consumer survey." Hear. Res. **14**, 8–20.

Keidser, G., Brew, C., Brewer, S., Dillon, H., Grant, F., and Storey, L. (**2005a**). "The preferred response slopes and two-channel compression ratios in twenty listening conditions by hearing-impaired and normal-hearing listeners and their relationship to the acoustic input," Int. J. Audiol. **44**, 656–670.

Keidser, G., Carter, L., and Dillon, H. (**2005b**). "The effect of advanced signal processing strategies in hearing aids on user performance and preference," in *Hearing Aid Fitting, Proceedings of the 21st Danavox Symposium*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen, pp. 307–320.

Keidser, G. (**1995**). "The relationship between listening conditions and alternative amplification schemes for multiple memory hearing aids," Ear Hear. **16**, 575–586.

Kiessling, J., Schubert, M., and Archut, A. (**1996**). "Adaptive fitting of hearing instruments by category loudness scaling (ScalAdapt)," Scand. Audiol. **25**, 153–160.

Kirkwood, B. C. (**2005**). "As good as the real thing? Performance differences for live versus recorded stimuli in an everyday listening task," in *Hearing Aid Fitting, Proceedings of the 21st Danavox Symposium*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen, pp. 519–529.

Kochkin, S. (**2007**). "MarkeTrak VII: Obstacles to adult non-user adoption of hearing aids," Hear. J. **60**, 31–34.

Lunner, T. (**2003**). "Cognitive function in relation to hearing aid use," Int. J. Audiol. **42**, S49–S58.

Lunner, T., Hellgren, J., Arlinger, S., and Elberling, C. (**1997**). "A digital filterbank hearing aid: improving a prescriptive fitting with subjective adjustments," Scand. Audiol. **26**, 169–176.

Magnusson, L., Karlsson, M., Ringdahl, A., and Israelsson, B. (**2001**). "Comparison of calculated, measured and self-assessed intelligibility of speech in noise for hearing-aid users," Scand. Audiol. **30**, 160–171.

Moore, B. C. J. (**2001**). "Dead regions in the cochlea: diagnosis, perceptional consequences, and implications for the fitting of hearing aids," Trends Amplif. **5**, 1–34.

Moore, B. C. J. (**2004**). "Dead regions in the cochlea: Conceptual foundations, diagnosis, and clinical applications," Ear Hear. **25**, 98–116.

Moore, B. C. J., Alcántara, J. I., and Glasberg, B. R. (**1998**). "Development and evaluation of a procedure for fitting multi-channel compression hearing aids," Br. J. Audiol. **32**, 177–195.

Moore, B. C. J., Marriage, J., Alcántara, J., and Glasberg, B. R. (**2005**).

"Comparison of two adaptive procedures for fitting a multi-channel compression hearing aid," Int. J. Audiol. **44**, 345–357.
Nordqvist, P., and Leijon, A. (**2004**). "An efficient, robust sound classification algorithm for hearing aids," J. Acoust. Soc. Am. **115**, 3033–3041.
Oticon. (**2007**). *Epoq*, p. 9. Oticon, publication 9101511000/05.07
Phonak. (**2006**). *Verve*, p. 8, 12. Phonak AG, publication 028-0326-02/v1.00/2006-05.
Pichora-Fuller, K. M., and Singh, G. (**2006**). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation," Trends Amplif. **10**, 29–59.
Punch, J. L., Robb, R., and Shovels, A. H. (**1994**). "Aided listener preferences in laboratory versus real-world environments," Ear Hear. **15**, 50–61.
Schweitzer, C., Mortz, M. S., and Vaughan, N. (**1999**). "Perhaps not by prescription, but by perception," High Performance Hearing Solutions **3**, 58–62.
Unitron (**2007**). *Yuu*, p. 3. Unitron hearing, publication 07-043 028-5256-02.
Zakis, J. A., McDermott, H. J., and Fisher, M. (**2004**). "An advanced sound processing system for hearing aid research," Iran. Audiol. **3**, 75–78.
Zakis, J. A., McDermott, H. J., and Dillon, H. (**2007**). "The design and evaluation of a hearing aid with trainable amplification parameters," Ear Hear. **28**, 812–830.

Effects of semantic predictability and regional dialect on vowel space reduction

Cynthia G. Clopper[a] and Janet B. Pierrehumbert
Department of Linguistics, Northwestern University, Evanston, Illinois 60208, USA

(Received 15 November 2006; revised 23 May 2008; accepted 8 June 2008)

This study explored the interaction between semantic predictability and regional dialect variation in an analysis of speech produced by college-aged female talkers from the Northern, Midland, and Southern dialects of American English. Previous research on the effects of semantic predictability has shown that vowels in high semantic predictability contexts are temporally and spectrally reduced compared to vowels in low semantic predictability contexts. In the current study, an analysis of vowel duration confirmed temporal reduction in the high predictability condition. An analysis of vowel formant structure and vowel space dispersion revealed overall spectral reduction for the Southern talkers. For the Northern talkers, more extreme Northern Cities shifting occurred in the high predictability condition than in the low predictability condition. No effects of semantic predictability were observed for the Midland talkers. These findings suggest an interaction between semantic and indexical factors in vowel reduction processes.
 [DOI: 10.1121/1.2953322]

PACS number(s): 43.70.Bk [CHS] Pages: 1682–1688

I. INTRODUCTION

Temporal and spectral vowel reduction has many sources, including lexical, semantic, and stylistic factors. At the lexical level, unstressed vowels are temporally and spectrally reduced relative to stressed vowels (Fourakis, 1991; Lindblom, 1963), vowels in high frequency words are temporally and spectrally reduced relative to vowels in low frequency words (Munson and Solomon, 2004), and vowels in words with few phonological neighbors (low density words) are temporally and spectrally reduced relative to vowels in words with many phonological neighbors (high density words; Munson and Solomon, 2004; Wright, 1997).

At the semantic level, predictable words are reduced relative to unpredictable words and repeated words are reduced relative to new words. Lieberman (1963) described the results of a series of studies in which talkers were recorded reading meaningful English sentences. The target word in each sentence was either highly predictable (e.g., nine in A stitch in time saves…) or unpredictable (e.g., nine in The number that you will hear is…). Lieberman (1963) found that the unpredictable words tended to be longer in duration, higher in amplitude, and pronounced with what he described as more precise articulation than the highly predictable words. As a result, when the words were excised from the sentence contexts and played back to naïve listeners, the unpredictable words were more intelligible overall than the predictable words. Similarly, Fowler and Housum (1987) found that words that had been used previously in an utterance were temporally reduced compared to words that were used for the first time. Like the unpredictable words in Lieberman's (1963) study, the "old" words in Fowler and Housum's (1987) study were also less intelligible when excised from running speech than the "new" words.

Jurafsky *et al.* (2001) analyzed a corpus of American English telephone conversations and found that function words were more likely to be reduced in both duration and vowel quality when the conditional probability of the target word given the previous word was high. For content words, they also found that duration was inversely correlated with the conditional probability of the target word given the preceding word. In addition, the relative frequency of the target word and the preceding word interacted, such that higher frequency targets were more reduced, whereas targets following high frequency words were less reduced.

Scarborough (2006) explicitly examined the relationship between lexical and semantic factors in vowel reduction and found an additive effect between semantic predictability and lexical neighborhood, such that "easy" words (high frequency words with few lexical neighbors) in high predictability semantic contexts were the most reduced, both spectrally and temporally, whereas "hard" words (low frequency words with many lexical neighbors) in low predictability semantic contexts were the least reduced.

With respect to speaking style, vowels in faster speech are spectrally reduced relative to vowels in slower speech (Fourakis, 1991) and vowels in plain or conversational speech are spectrally and temporally reduced relative to vowels in clear speech (Lindblom, 1990; Picheny *et al.* 1986). Taken together, the lexical, semantic, and stylistic effects on vowel reduction suggest a model of production in which the phonetic targets for individual vowels are hyperarticulated relative to what is typically produced (Johnson *et al.* 1993). According to this account, talkers selectively reduce vowels based on current processing constraints in production and potential listener-oriented constraints, such as noise, native language background, or hearing impairment (Lindblom,

[a]Present address: Department of Linguistics, Ohio State University, Columbus, OH 43210.

1990). Thus, high frequency and low density words, semantically predictable words, and casually produced speech all exhibit vowel reduction relative to low frequency and high density words, semantically unpredictable words, and carefully produced speech.

A. Vowel reduction and dialect variation

Several recent studies have obtained evidence for an interaction between vowel reduction processes, such as those described above, and indexical sources of variability. For example, Scarborough (2006) observed a greater effect of lexical neighborhood density on spectral reduction for /o/ than for the other vowels she measured. One possible explanation for her finding is that her talkers were from California and exhibited fronted /o/'s in their speech. It is well documented in the sociolinguistic literature that participants produce more dialect-specific variants in less formal settings, such as interview speech, than in more formal settings, such as reading passages or word lists (e.g., Labov, 1972). This effect of formality on dialect variation is similar to the effect of clear versus conversational speaking styles on vowel reduction, in that both effects may reflect a listener-oriented attempt on the part of the speaker to better approximate a set of hyperarticulated and/or standard variants. Thus, Scarborough's (2006) density × vowel interaction may reflect the effects of both dialect-specific /o/ fronting and lexically induced spectral reduction of the /o/ in low density words.

Similarly, Munson (2007b) observed an interaction between lexical density and gender typicality for the male talkers in his study of vowel space reduction. The effect of gender typicality on vowel space reduction was significant for the male talkers for the easier, low density words, but not for the harder, high density words. Together, Scarborough's (2006) and Munson's (2007b) results suggest that greater indexical variability may be observed under the same conditions in which vowel reduction is typically observed. However, Jacewicz *et al.* (2006) reported more extreme dialect variants in prosodically strong positions compared to prosodically weak positions, suggesting that some nonreduced contexts may also favor indexical marking.

The current study was designed to explore the relationship between the effects of dialect variation and semantic predictability on vowel reduction. Specifically, we examined the degree of Northern Cities vowel shifting across two semantic contexts: high predictability and low predictability. The Northern Cities Chain Shift (NCCS) involves the fronting and raising of /æ/, the lowering and fronting of /ɑ/, the lowering and fronting of /ɔ/, the backing and/or lowering of /ɛ/, the backing of /ʌ/, and the backing of /ɪ/ (Labov, 1998). A schematic of the NCCS is shown in Fig. 1. The NCCS is typically adopted in sequential order, beginning with the raising and fronting of /æ/, as indicated by the numbers in Fig. 1 (Labov *et al.*, 2006). It should also be noted that for some of the vowels, the NCCS and overall vowel reduction processes pull the vowels in different directions in the F1 × F2 space. For example, both the NCCS and vowel reduction can lead to raising of /æ/, but the NCCS produces /æ/ fronting and vowel reduction produces /æ/ backing. Thus, the interaction between the two processes for /æ/ should be reflected along the F2 dimension more than the F1 dimension. The acoustic characteristics and sequential adoption of the NCCS allowed us to explore differences in dialect-specific variants across the two semantic predictability conditions, as well as how the direction of the vowel shifts in the acoustic space affected the interaction between dialect variation and semantic predictability, and how this interaction affected early-adopted vowels in the NCCS compared to late-adopted vowels.

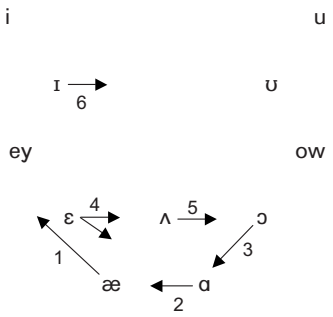


FIG. 1. Schematic of the Northern Cities Chain Shift. Numbers indicate the order in which each component is adopted.

II. METHODS

A. Talkers

Forty female talkers were selected from the Nationwide Speech Project corpus (Clopper and Pisoni, 2006) and the Indiana Speech Project corpus (Clopper *et al.*, 2002). The talkers were all 18–25 year old white native speakers of American English. The talkers were classified into three broad geographical dialects based on their region of origin: Southern, Northern, and Midland. The Southern group included eight women from the Louisville, KY metropolitan area and two women from Texas, for a total of ten Southerners. The Northern group was composed of eight women from the Chicago metropolitan area, and seven women from Northern Indiana: two from South Bend and five from Fort Wayne. Finally, the Midland group included eight women from Central Indiana and seven women from the Evansville area in Southern Indiana. The talkers were selected for recording and assigned to the dialect groups based on residential history. Each of the women had lived in the same dialect region until at least age 18 and both of her parents were also from the same region. Previous acoustic analyses of the vowel productions of a subset of the talkers confirmed dialect-specific differences across the three groups (Clopper *et al.*, 2005). Male talkers were excluded from the study because recordings were available for only five male talkers from each region which did not allow for adequate statistical power for a comparison across gender.

B. Stimulus materials

Four target words for each of four vowels were selected for comparison in two semantic contexts, for a total of 32 tokens per talker. The four vowels were /æ/ and /ɑ/, two of the earliest-adopted vowels in the Northern Cities Chain

Shift, /ʌ/, one of the later-adopted vowels in the NCCS, and /i/, a vowel not involved in the NCCS (see Fig. 1; Labov, 1998). Each of the target words was produced by each of the talkers in sentence-final position in two semantic contexts: high predictability and low predictability. All of the sentences were taken from the Speech Perception in Noise (SPIN) test of Kalikow *et al.* (1977). The high predictability sentences were five to eight words long and were constructed such that the final word in the sentence was predictable from the preceding semantic context. Semantic predictability was confirmed by Kalikow *et al.* (1977) in the development of the SPIN sentences using a sentence completion task. The low predictability sentences were also short meaningful sentences, but the final words were not predictable from the preceding semantic context. The final target words were produced with a pitch accent by all of the talkers. Each target word examined in this study was, therefore, produced in both a highly predictable and an unpredictable context. A complete list of the stimulus materials is shown in the Appendix.

The sentences were recorded as part of a larger corpus that included isolated words, sentences, passages, and interview speech (Clopper *et al.*, 2002; Clopper and Pisoni, 2006). The 16 high predictability sentences in the current study were selected from a larger set of high predictability sentences produced by each talker. The 16 low predictability sentences were selected from a larger set of low predictability sentences produced by each talker. Due to the design of the original corpora, only four utterances per vowel and only one repetition per utterance were available for analysis in the current experiment. All of the materials were digitally recorded directly to a Macintosh Powerbook G3 laptop at a sampling rate of 44.1 kHz and 16 bit resolution in a sound-attenuated booth using a Shure head-mounted microphone (SM10A).

C. Procedure

Three acoustic measurements were obtained from each of the tokens using spectral and waveform views in Praat: vowel duration in milliseconds, first formant frequency (F1) in hertz, and second formant frequency (F2) in hertz. The duration measurements were obtained by hand by the first author (C.G.C.). The formant frequency measurements were obtained at the first-third temporal point, defined as the onset of the vowel plus one-third of the vowel duration, in order to capture the spectral "nucleus" of each vowel. Given that /æ/ is often diphthongal for Northern talkers and /i/ is often diphthongal for Southern talkers, the first-third temporal point was selected to ensure that the measurements reflected the nucleus of each vowel (see Clopper *et al.*, 2005). Formant values were extracted by hand using Praat's standard formant tracking tool with a 25 ms window and a 12th-order LPC analysis. For the small number of stimulus materials for which the 12th-order LPC analysis failed to correctly track the formants, a 10th- or 14th-order LPC analysis was used to extract the formant frequencies. The formant frequency measures were converted to the Bark scale for analysis (Traunmüller, 1990).

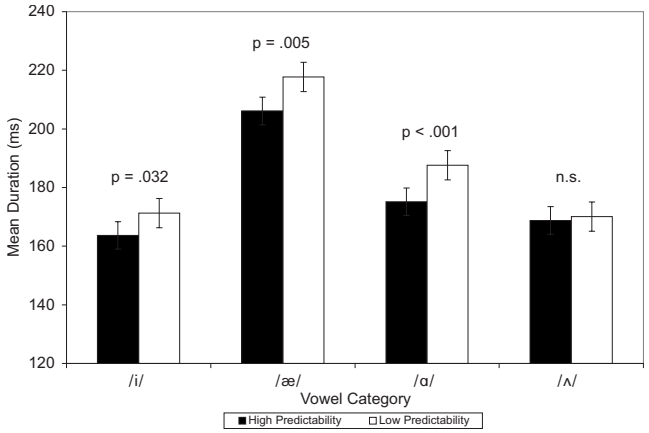


FIG. 2. Mean vowel duration in milliseconds for each vowel in each semantic predictability condition, collapsed across target word and talker dialect. Error bars are standard error.

III. RESULTS

A summary of the duration results is shown in Fig. 2 for each vowel in each semantic predictability condition. For each vowel, a repeated measures analysis of variance (ANOVA) with semantic predictability [high predictability (HP) or low predictability (LP)] as a within subjects factor and talker dialect (North, Midland, or South) as a between subjects factor was conducted. Significant main effects of semantic predictability were obtained for /i/ [$F(1,37)=5.0$, $p=0.032$], /æ/ [$F(1,37)=8.9$, $p=0.005$], and /ɑ/ [$F(1,37)=17.2$, $p<0.001$]. In all three cases, the vowels in the low predictability condition were longer than the vowels in the high predictability condition. The effect of semantic predictability did not reach significance for /ʌ/. Talker dialect was not a significant factor for any of the vowels and none of the semantic predictability × talker dialect interactions were significant. These results confirm an effect of semantic predictability on vowel duration for three of the four vowels across all three talker dialects.

A summary of the spectral analysis is shown in Fig. 3. The filled symbols represent the mean first and second for-

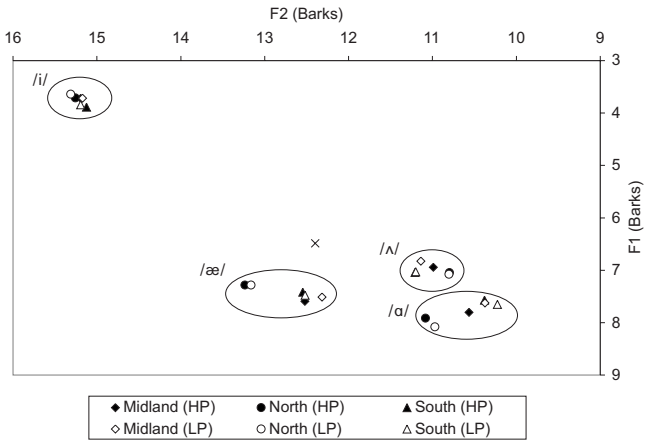


FIG. 3. Mean first and second formant frequencies of /i/, /æ/, /ɑ/, and /ʌ/ in high predictability (filled) and low predictability (open) contexts for the Midland (diamond), Northern (circle), and Southern (triangle) talkers. The grand mean of the vowel space across all talkers and conditions is indicated by the X.

mant frequencies for each of the vowels in the high predictability condition. The open symbols represent the mean F1 and F2 values for the low predictability condition. The grand mean of the vowel space, collapsed across all talkers and semantic predictability conditions, is indicated by the X to provide an indication of the predicted direction of spectral vowel reduction.

An inspection of Fig. 3 suggests that the Northern talkers produced fronted /æ/'s and /ɑ/'s and backed /ʌ/'s compared to the Midland and Southern talkers. Separate repeated measures ANOVAs were conducted on the F1 and F2 measures for each vowel with semantic predictability (HP or LP) as a within subjects factor and talker dialect (North, Midland, or South) as a between subjects factor. The main effect of dialect was significant for the F2 of /æ/ [$F(2,37)=16.4$, $p<0.001$] and the F2 of /ɑ/ [$F(2,37)=10.9$, $p<0.001$]. *Post hoc* Tukey tests revealed that the Northern talkers produced significantly fronted /æ/'s and /ɑ/'s compared to the Midland or Southern talkers (all $p<0.005$), reflecting Northern Cities shifted productions of /æ/ and /ɑ/ for the Northern talkers. The main effect of dialect for the F2 of /ʌ/ did not reach significance.

The series of repeated measures ANOVAs on formant frequency values also revealed a significant main effect of semantic predictability for the F1 of /ɑ/ [$F(1,37)=6.4$, $p=0.015$] and the F2 of /ɑ/ [$F(1,37)=11.1$, $p=0.002$], as well as a significant semantic predictability × dialect interaction for the F1 of /ɑ/ [$F(2,37)=4.0$, $p=0.028$]. None of the other main effects or interactions were significant. For the F1 of /ɑ/, the low predictability variants were produced with a larger F1 than the high predictability variants. For the F2 of /ɑ/, the low predictability variants were produced with a smaller F2 than the high predictability variants. Thus, in both the F1 and F2 dimensions, significant reduction was observed for /ɑ/ in the high predictability condition relative to the low predictability condition. An inspection of Fig. 3 reveals the locus of the semantic predictability × dialect interaction for the F1 of /ɑ/. While the Northern and Southern talkers both produced lower /ɑ/'s in the low predictability condition, the Midland talkers produced lower /ɑ/'s in the high predictability condition. Thus, the reduction of /ɑ/ in the F1 dimension in the high predictability condition was not observed consistently across dialects; unlike the Northern and Southern talkers, the Midland talkers did not exhibit F1 reduction of /ɑ/ in the high predictability context.

Significant effects of semantic predictability were observed for only the F1 and F2 of /ɑ/. However, vowel space reduction and expansion are typically assessed in a two-dimensional space and the analysis of first and second formant frequencies examined each dimension separately. A vowel space dispersion analysis was therefore conducted to explore reduction in the two-dimensional F1 × F2 vowel space. Dispersion was calculated separately for each vowel in each semantic predictability context and was defined as the Euclidean distance in the F1 × F2 Bark space from the target vowel to the average F1 and F2 across all of the vowels analyzed (see Wright, 1997). Overall vowel space dispersion for each semantic predictability condition was defined as the mean of the four individual vowel dispersion measures.

TABLE I. Mean overall vowel space dispersion in Barks for each of the three dialects in the high predictability and low predictability conditions. Standard deviations are shown in parentheses.

Dialect	High predictability	Low predictability	Difference (LP−HP)
Midland	2.19 (0.13)	2.14 (0.13)	−0.05
North	2.24 (0.21)	2.28 (0.20)	0.04
South	2.08 (0.16)	2.15 (0.12)	0.07

The mean overall vowel space dispersion measures for each of the dialects in the high and low predictability conditions are shown in Table I. A repeated measures ANOVA on vowel space dispersion with semantic predictability (HP or LP) as a within subjects factor and talker dialect (North, Midland, or South) as a between subjects factor revealed a significant semantic predictability × dialect interaction [$F(2,37)=5.1$, $p=0.011$]. Neither of the main effects were significant. *Post hoc* paired comparison *t*-tests on vowel space dispersion in each semantic predictability condition for each dialect revealed the locus of the interaction. A significant difference in vowel space dispersion as a result of semantic predictability was observed for the Southern talkers [$t(9)=-2.8$, $p=0.020$]. Semantic predictability did not significantly affect overall vowel space dispersion for the Midland or the Northern talkers.

The significant semantic predictability × dialect interaction in the vowel space dispersion analysis may reflect differences in spectral reduction between the four different vowels included in the analysis. Specifically, the overall measure of vowel dispersion may have been reduced for the Northern talkers because some vowels were spectrally reduced due to semantic predictability while others were spectrally dispersed due to more extreme Northern Cities shifting. To further explore this difference in overall vowel dispersion effects across the three dialects, the individual dispersion measures for each vowel were analyzed. The mean difference in dispersion for each of the vowels between the low predictability and the high predictability conditions is shown in Fig. 4. Positive differences indicate greater dispersion in the low predictability condition than the high predictability condition. Negative differences indicate greater dispersion in the high predictability condition than the low predictability condition.

The Southern talkers, who exhibited an effect of semantic predictability on overall vowel space dispersion, produced /i, æ, ɑ/ with greater dispersion in the low predictability condition than the high predictability condition, but /ʌ/ with less dispersion in the low predictability condition than the high predictability condition. The Northern talkers produced /i, ɑ/ with greater dispersion and /ʌ/ with less dispersion in the low predictability condition than the high predictability condition, similar to the Southern talkers. However, the Northern talkers also produced /æ/ with less dispersion in the low predictability condition than the high predictability condition, unlike the Southern talkers. For the Midland talk-

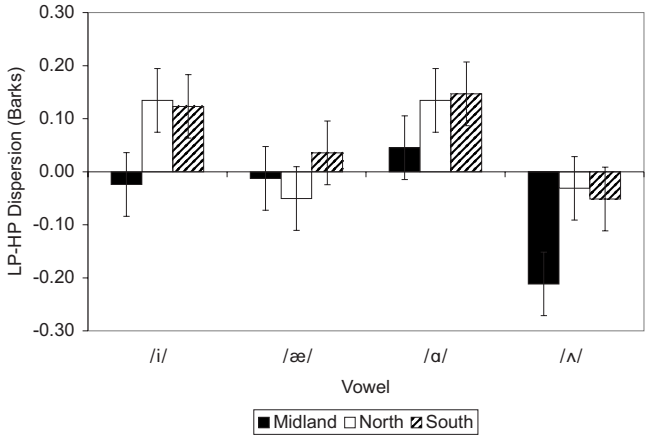


FIG. 4. Mean dispersion difference between the low and high predictability conditions for each vowel for each dialect. Error bars are standard error.

ers, the difference scores for the individual vowels revealed very small differences across the two semantic predictability conditions, with the exception of /ʌ/, which was produced with less dispersion in the low predictability condition than the high predictability condition. A repeated measures ANOVA on the dispersion difference scores with vowel (/ i, æ, ɑ, ʌ/) as a within subject factor and dialect as a between subject factor revealed a significant vowel × dialect interaction [$F(6,111)=7.2$, $p<0.001$]. The main effects of vowel and dialect were not significant. An inspection of Fig. 4 suggests that the Northern dispersion difference for /æ/ is the locus of this interaction: for the Northern talkers, greater dispersion is observed for /æ/ in the high predictability context than the low predictability context. High predictability /æ/'s were more fronted than low predictability /æ/'s for the Northern talkers (see Fig. 3).

IV. DISCUSSION

The duration analysis revealed a significant main effect of semantic predictability for all three dialect groups for three of the four vowels examined. For /i, æ, ɑ/, the vowels were shorter in the high predictability condition than the low predictability condition. In addition, the spectral analysis revealed a significant main effect of semantic predictability for the F1 and F2 of /ɑ/. High predictability /ɑ/'s were raised and fronted relative to low predictability /ɑ/'s for the Northern and Southern talkers, and were fronted, but not raised, relative to the low predictability /ɑ/'s for the Midland talkers. These results are consistent with previous research on the effects of semantic predictability on vowel reduction, which has revealed temporal and spectral reduction in high predictability conditions compared to low predictability conditions (Lieberman, 1963; Scarborough, 2006). The spectral analysis also revealed a significant main effect of talker dialect for the F2 of /æ/ and /ɑ/. The Northern talkers produced more fronted /æ/'s and /ɑ/'s than the Midland or Southern talkers, consistent with the early stages of the Northern Cities Chain Shift (Labov, 1998).

An interaction between semantic predictability and talker dialect was observed in the vowel space dispersion analysis. The effect of semantic predictability on overall vowel space dispersion was observed for the Southern talkers but not the Northern or Midland talkers. The Southern vowels exhibited less dispersion overall in the high predictability condition than the low predictability condition, consistent with previous research on the effects of semantic predictability on vowel reduction (Lieberman, 1963; Scarborough, 2006).

For the Northern talkers, spectral reduction of /i/ and /ɑ/ was accompanied by greater fronting of /æ/ in the high predictability condition (see Fig. 3), leading to the significant vowel × dialect interaction in the individual vowel dispersion analysis and diminishing the effect of semantic context for the Northern talkers in the overall vowel space dispersion analysis. The greater fronting of /æ/ reflects a more extreme production of the Northern Cities shift. These results suggest that vowel production is affected by both semantic constraints, such as predictability, and indexical constraints, such as regional dialect, and that these factors can interact. In particular, the effect of semantic predictability on overall vowel space dispersion is apparently diminished for the Northern talkers due to more extreme Northern Cities /æ/ fronting in the high predictability condition.

This interpretation of the current results is consistent with the findings reported by Munson (2007b) and Scarborough (2006), who observed interactions between lexical constraints, such as word frequency and neighborhood density, and indexical constraints, such as gender prototypicality and regional dialect, on vowel reduction. More extreme indexical variants are produced in the easy processing conditions (e.g., high predictability sentences or low density words) than in the hard processing conditions. These findings suggest that listeners can represent multiple acoustic-phonetic targets for a given vowel category. Specifically, the relatively educated participants in the current study were exposed to local or regional vowel variants (e.g., Northern or Southern) and supraregional standard variants (e.g., General American). For the Northern talkers, the local and standard variants of /æ/ were in competition for production, and the more extreme local variant was produced in the high predictability context relative to the low predictability context. For the Southern talkers, the local and standard variants of the vowels examined were not in competition because both variants of each vowel were phonetically similar, and traditional effects of semantic predictability on vowel space reduction were observed.

The representation of multiple targets for a given vowel category, and competition among those targets in processing, can be accounted for by a class of models, including exemplar theory, in which talkers learn probability distributions of phonemes with some degree of phonetic detail (Pierrehumbert, 2006). These models would predict that the talkers in the current study maintain simultaneous representations of local and standard vowel variants and that those variants are differentially activated under different conditions. Hay *et al.*, (2006) proposed a similar account for the perception of an ongoing sound change in New Zealand, in which listeners maintain multiple acoustic-phonetic representations of a vowel category as it merges with another category.

The activation of the local versus the standard variant across the two semantic predictability conditions may be driven by at least two processing considerations. One possible account of the current results is that the increase in Northern Cities shifting in the high predictability condition reflects a listener-oriented adaptation on the part of the talker. Specifically, talkers choose more shifted variants in the high predictability condition because they know that the listener will be able to access the word easily based on the semantic context. The talker can therefore afford to index social information to the listener without significantly affecting intelligibility. Alternatively, the current results may reflect talker-oriented processing differences in high versus low predictability sentences. For example, the temporal reduction that was observed for all of the talkers may reflect faster lexical retrieval or phonetic implementation processes in the high predictability condition relative to the low predictability condition (Munson, 2007a). In speech perception, some sources of indexical information are processed more slowly than linguistic information and therefore may not affect performance in tasks in which speeded responses are required (McLennan and Luce, 2005). Similar differences in processing lexical, semantic, and indexical information in speech production may account for the interactions in the current study between semantic and indexical factors in vowel reduction. Additional research is needed to explore the effects of semantic predictability and talker dialect on listener-directed speech (such as clear speech) and on speech produced in speeded and nonspeeded production tasks to distinguish between these two accounts of the current data.

Two components of the results require additional comment. First, /ʌ/ did not exhibit either temporal or spectral reduction in the high predictability condition for any of the three dialects. This finding is consistent with Wright's (1997) observation that the lax vowels /ɪ, ɛ, ʌ/ exhibit little to no reduction in low density words compared to high density words, whereas the tense vowels /i, æ, ɑ, ɔ, u/ show greater dispersion in high density words than low density words. Thus, the lack of reduction of /ʌ/ in the current study may reflect more general differences in reduction processes for tense and lax vowels in English.

Second, for the Midland talkers, the effect of semantic predictability on overall vowel space dispersion was not observed. In addition, whereas the Northern and Southern talkers produced raised /ɑ/'s in the high predictability condition, the Midland talkers produced lowered /ɑ/'s in the high predictability condition relative to the low predictability condition, leading to the significant semantic predictability × dialect interaction in the analysis of the F1 of /ɑ/. Thus, in the high predictability condition, the Midland talkers exhibited temporal reduction to the same extent as the Northern and Southern talkers, but they did not produce spectrally reduced vowels and, in the case of /ʌ/, exhibited greater dispersion in the high predictability condition than the low predictability condition. An inspection of the individual data revealed that the dispersion patterns were not due to a single outlier or a geographical subset of the Midland talkers but were observed across the entire group. These findings for the Midland talkers reflect a failure to replicate previous work on the effects of semantic predictability on vowel space reduction (Lieberman, 1963; Scarborough, 2006) and present a challenge to the probabilistic models that can account for the data obtained from the Northern and Southern talkers. However, we know of no other model that would be able to account for this set of findings, so this set of results for the Midland talkers remains a challenge to any account of the effects of linguistic and indexical factors on vowel space dispersion.

Taken together, the results of the current study confirm previously observed significant effects of semantic predictability and regional dialect on vowel production. In addition, significant interactions between semantic predictability and talker dialect were observed, consistent with the results of a growing literature exploring the interactions between lexical and indexical factors in speech production and perception and with a class of probabilistically based models of language processing in which phonological representations are phonetically detailed. Finally, vowel-specific findings suggest that the effects of semantic predictability may be limited to certain classes of vowels (e.g., tense vowels in English).

ACKNOWLEDGMENTS

This work was supported by NIH NRSA Postdoctoral Fellowship F32 DC007237 and James S. McDonnell Foundation Award No. 21002061 to Northwestern University.

APPENDIX

Vowel	High predictability	Low predictability
/i/	The chicken pecked corn with its beak.	She is glad Bill called about the beak.
	The wedding banquet was a feast.	We could consider the feast.
	Watermelons have lots of seeds.	You have considered the seeds.
	Ruth had a necklace of glass beads.	Tom has been discussing the beads.
/æ/	Please wipe your feet on the mat.	Peter has considered the mat.
	Paul hit the water with a splash.	Bob has discussed the splash.
	The cow gave birth to a calf.	She hopes Jane called about the calf.
	The cut on his knee formed a scab.	The boy would discuss the scab.
/ɑ/	Paul was arrested by the cops.	Ruth hopes Bill called about the cop.
	The shepherds guarded their flock.	Paul should have discussed the flock.

Vowel	High predictability	Low predictability
	Tighten the belt by a notch.	The woman considered the notch.
	The plow was pulled by an ox.	The man should discuss the ox.
/ʌ/	The bird of peace is the dove.	Peter could consider the dove.
	The heavy rains caused a flood.	The class should consider the flood.
	Cut the meat into small chunks.	I did not know about the chunks.
	Paul took a bath in the tub.	Miss Smith knows about the tub.

Clopper, C. G., Carter, A. K., Dillon, C. M., Hernandez, L. R., Pisoni, D. B., Clarke, C. M., Harnsberger, J. D., and Herman, R. (**2002**). "The Indiana Speech Project: An overview of the development of a multi-talker multi-dialect speech corpus," Research on Spoken Language Processing Progress Report No. 25, Speech Research Laboratory, Indiana University, Bloomington, pp. 367–380.

Clopper, C. G., and Pisoni, D. B. (**2006**). "The Nationwide Speech Project: A new corpus of American English dialects," Speech Commun. **48**, 633–644.

Clopper, C. G., Pisoni, D. B., and de Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.

Fourakis, M. (**1991**). "Tempo, stress, and vowel reduction in American English," J. Acoust. Soc. Am. **90**, 1816–1827.

Fowler, C. A., and Housum, J. (**1987**). "Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction," J. Mem. Lang. **26**, 489–504.

Hay, J., Warren, P., and Drager, K. (**2006**). "Factors influencing speech perception in the context of a merger-in-progress," J. Phonetics **34**, 458–484.

Jacewicz, E., Salmons, J., and Fox, R. A. (**2006**). "Cross-dialectal differences in the effects of prosody on vowels and chain shifting," paper presented at New Ways of Analyzing Variation 35, Columbus, OH, November.

Johnson, K., Flemming, E., and Wright, R. (**1993**). "The hyperspace effect: Phonetic targets are hyperarticulated," Language **83**, 505–528.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (Benjamins, Amsterdam), pp. 229–254.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Labov, W. (**1972**). *Sociolinguistic Patterns* (University of Pennsylvania, Philadelphia).

Labov, W. (**1998**). "The three dialects of English," in *Handbook of Dialects and Language Variation*, edited by M. D. Linn (Academic, San Diego), pp. 39–81.

Labov, W., Ash, S., and Boberg, C. (**2006**). *Atlas of North American English* (Mouton de Gruyter, New York).

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.

Lindblom, B. E. F. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B. E. F. (**1990**). "Explaining phonetic variation: A sketch of the H&H Theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Amsterdam), pp. 403–439.

McLennan, C. T., and Luce, P. A. (**2005**). "Examining the time course of indexical specificity effects in spoken word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **31**, 306–321.

Munson, B. (**2007a**). "Lexical access, lexical representation, and vowel production," in *Papers in Laboratory Phonology IX*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, Berlin), pp. 201–228.

Munson, B. (**2007b**). "Lexical characteristics mediate the influence of sex and sex typicality on vowel-space size," Proceedings of the 16th International Congress of Phonetic Sciences, pp. 885–888.

Munson, B., and Solomon, N. P. (**2004**). "The effect of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–445.

Pierrehumbert, J. (**2006**). "The next toolkit," J. Phonetics **34**, 516–530.

Scarborough, R. (**2006**). "Lexical and contextual predictability: Confluent effects on the production of vowels," paper presented at the tenth Laboratory Phonology Conference, Paris, France, June.

Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.

Wright, R. (**1997**). "Lexical competition and reduction in speech: A preliminary report," Research on Spoken Language Processing Progress Report No. 21, Speech Research Laboratory, Indiana University, Bloomington, pp. 471–485.

Influence of flow separation location on phonation onset[a)]

Zhaoyan Zhang[b)]
UCLA School of Medicine, 31-24 Rehabilitation Center, 1000 Veteran Avenue, Los Angeles, California 90095-1794

(Received 19 January 2008; revised 17 June 2008; accepted 18 June 2008)

The influence of flow separation location on eigenmode synchronization and phonation onset was investigated in a two-dimensional, aeroelastic, continuum model of phonation. A linear stability analysis showed that flow separation played a critical role in initiating eigenmode synchronization and phonation. For a given glottal configuration, a small variation in the flow separation location along the vocal fold surface may lead to a qualitatively different eigenmode-synchronization pattern, and different phonation threshold pressure and frequency. This high sensitivity suggests a need for phonation models to be capable of accurate prediction of the flow separation location. Analysis with different glottal channel geometries showed that a minimum phonation threshold pressure existed for a rectangular glottal channel, consistent with previous experiments. However, in contrast to previous theoretical analyses, this study showed that phonation was facilitated, rather than prohibited, by the upstream movement of the flow separation point within a divergent glottis.
 [DOI: 10.1121/1.2957938]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 1689–1694

I. INTRODUCTION

A recent theoretical analysis (Zhang *et al.*, 2007) showed that phonation onset occurred as two eigenmodes of the vocal fold synchronized due to a cross-mode coupling effect of the glottal flow. The details of the eigenmode synchronization process and the condition at which phonation onset occurs depend on both vocal fold biomechanics and the ability of the glottal flow to synchronize two or more structural eigenmodes. The biomechanical properties of the vocal folds determine the natural frequencies of the vocal fold structure, the initial state of the coupled system in the eigenmode synchronization process. The strength of the cross-mode coupling effect of the glottal flow depends on the intraglottal pressure distribution and how it varies with the vocal fold surface motion. It is generally accepted that airflow through the glottis remains laminar until the flow separates from the vocal fold surface. Downstream of the flow separation point, a jet forms, which eventually transitions into a turbulent flow, accompanied by a slight pressure recovery toward the supraglottal pressure. The exact flow separation location therefore determines the intraglottal pressure distribution and, for a given glottal configuration, the nature and strength of the eigenmode-synchronizing effect of the glottal flow.

The flow separation location is highly dependent on both the vocal fold geometry and the glottal flow field. Studies on steady flow through orifices of various configurations (Scherer *et al.*, 2001; Zhang *et al.*, 2004; Kucinschi *et al.*, 2006) showed that airflow separated around the superior edge of the medial surface for a convergent or straight glottis. For divergent glottis, the flow separation location moved upstream to a location inside the glottis. With increasing Reynolds number, the flow separation point is generally expected to move further upstream along the divergent medial surface. During phonation, the glottis changes alternately from convergent, straight, to divergent. The flow separation location may vary accordingly during one cycle of vocal fold oscillation. For example, numerical simulations on unsteady flow through forced-oscillating glottis (Zhao *et al.*, 2002; Alipour and Scherer, 2004) showed that the flow separation point moved upstream when the glottal divergence angle exceeded a certain threshold value, and persisted upstream into the convergent phase of the cycle. Recent numerical simulation (Sciamarella and Le Quere, 2008) also showed that the flow separation process during phonation was highly unsteady and the vocal fold motion had a significant influence on the flow separation location.

Accurate prediction of the flow separation location requires solving the full three-dimensional flow equations (e.g., Zhao *et al.*, 2002), and is computationally expensive. For practical applications, simplified models are often used. Early models of phonation generally assumed the airflow to separate from the glottal wall at a fixed location. For example, the two-mass model (Ishizaka and Flanagan, 1972; Ishizaka, 1981) assumed flow separation occurred at the sharp edge of the superior medial surface. Pelorson *et al.* (1994) proposed a quasisteady flow separation model in which the flow separation location was predicted based on boundary layer theory. The flow separation point was allowed to move upstream when the glottis became divergent. A separation constant was also used in many models, which assume flow separation to occur at a point downstream of the minimum glottal constriction with a cross-sectional area equal to the minimum glottal area multiplied by the separation constant [e.g., 1.2, see Lous *et al.* (1998)]. Although these simplifications greatly reduce the computational costs,

[a)]Portions of this work were presented at the 153rd Meeting of the Acoustical Society of America, Salt Lake City, Utah, June 2007.
[b)]Electronic mail: zyzhang@ucla.edu

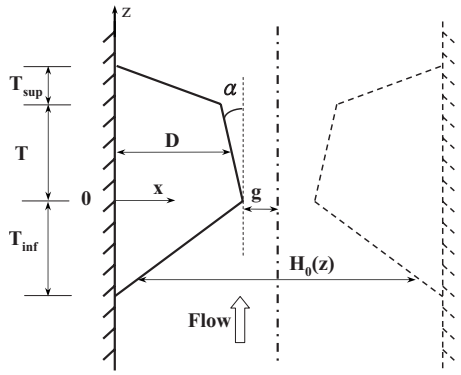


FIG. 1. Two-dimensional vocal fold model and the glottal channel. The flow direction is along the positive z axis. The coupled vocal fold-flow system was assumed to be symmetric about the glottal channel centerline so that only the left-hand half of the system was considered in this study. D is the vocal fold depth in the medial-lateral direction at the midpoint of the medial surface; T, T_{inf}, and T_{suf} are the thicknesses of the entrance, medial surface, and exit sections of the vocal fold in the flow direction, respectively; H_0 is the prephonatory glottal channel width; g is the minimum prephonatory glottal half-width of the glottal channel; The divergence angle α is the angle formed by the medial surface of the vocal fold with the z axis.

it is unclear how they might affect the predicted phonation characteristics. It would be useful to understand the sensitivity of phonation to slight variation in the flow separation location. In other words, how accurate does a prediction of the flow separation location have to be to reasonably capture the major dynamics of the fluid–structure interaction?

The study is a sensitivity study, with the objective of investigating how changes in the flow separation location affect eigenmode synchronization and the resulting phonation onset. A two-dimensional self-oscillating continuum model of phonation as in Zhang *et al.* (2007) was used. Continuum models of phonation allow more realistic reproduction of the complex fluid–structure interaction within the glottis. As in that study, phonation onset characteristics were calculated from a linear stability analysis. We will show that flow separation plays a central role in initiating phonation, and, for a given glottal geometry, a slight variation of the flow separation point may have a significant influence on the resulting phonation onset characteristics. We will also show that, in contrast to previous understanding, phonation is facilitated, rather than prohibited [as discussed by Lucero (1998), see below], by the upstream movement of the flow separation point within a divergent glottis.

II. MODEL DESCRIPTION

Consider a continuum vocal fold model coupled with a one-dimensional potential flow, which separates at a certain location along the vocal fold surface (Fig. 1). The vocal folds were modeled as a two-dimensional, isotropic, plane strain, elastic layer of Young's modulus $\bar{E}$ and density $\bar{\rho}_{vf}$. No vocal tract was considered and the flow rate at the glottal entrance was assumed to be constant. For simplicity, left–right symmetry in the flow and vocal fold vibration about the glottal channel centerline was imposed and only half the system was considered. A nondimensional formulation of system equations was used. The vocal fold thickness $\bar{T}$, vocal fold density $\bar{\rho}_{vf}$, and the wave velocity of the vocal fold structure $\sqrt{\bar{E}/\bar{\rho}_{vf}}$ were used as the reference length, density, and velocity scales, respectively. The nondimensional variables are defined as follows:

$$T = 1, \quad \rho_{vf} = 1, \quad E = 1,$$

$$D = \bar{D}/\bar{T}, \quad g = \bar{g}/\bar{T}, \quad T_{\text{inf}} = \bar{T}_{\text{inf}}/\bar{T}, \quad T_{\text{sup}} = \bar{T}_{\text{sup}}/\bar{T},$$

$$\rho_{\text{f}} = \bar{\rho}_f/\bar{\rho}_{vf}, \quad P_s = \bar{P}_s/\bar{E},$$

$$U_j = \bar{U}_j/\sqrt{\bar{E}/\bar{\rho}_{vf}}, \quad f = \bar{f}/(\sqrt{\bar{E}/\bar{\rho}_{vf}}/\bar{T}), \tag{1}$$

where ρ_f is the density of air, U_j is the mean jet velocity, P_s is the mean subglottal pressure, and f is phonation frequency. Symbols without overbars denote nondimensional variables. The physical values can be recovered by multiplying nondimensional values by the corresponding reference scales. All variables in the following are in nondimensional forms unless otherwise stated.

The linear stability analysis as in Zhang *et al.* (2007) was applied to the coupled fluid–structure system, from which the phonation threshold pressure and frequency and vocal fold vibration characteristics at phonation onset were calculated. A brief description of the linear stability analysis procedure is given in this section. Readers are referred to the original paper for a detailed derivation of system equations. The system equations were derived from Lagrange's equations (Zhang *et al.*, 2007):

$$(M - Q_2)\ddot{q} + (C - Q_1)\dot{q} + (K - Q_0)q = 0, \tag{2}$$

where q is the generalized coordinate vector. The three matrices M, C, and K represent the mass, damping, and stiffness matrices of the vocal fold structure, respectively. A proportional structural damping was assumed for the vocal fold material so that the structural mass and damping matrices were related by $C = \sigma\omega M$, where σ is the constant structural loss factor and ω is the angular frequency. For human vocal folds, the loss factor is in the range of 0.4–1.0 at low frequencies and 0.2–0.6 at high frequencies (10–15 Hz) (Chan and Titze, 1999). The term $Q_2\ddot{q} + Q_1\dot{q} + Q_0 q$ in Eq. (2) is the generalized force associated with the fluctuating flow pressure along the vocal fold surface induced by vocal fold vibration. The fluctuating flow pressure was obtained as follows. First, Bernoulli's equation and the continuity equation of airflow were linearized around the mean state of the coupled airflow–vocal fold system (Zhang *et al.*, 2007). As boundary conditions for the flow domain, a zero fluctuating flow velocity was imposed at the glottal entrance and a zero fluctuating pressure at the flow separation location, assuming a zero pressure recovery downstream of the flow separation location. The fluctuating pressure was then obtained by solving the linearized flow equations with these boundary conditions. The flow pressure can be decomposed into a flow-induced stiffness term, which is proportional to vocal fold displacement and represented by matrix Q_0, a flow-induced damping term, which is proportional to vocal fold velocity and represented by matrix Q_1, and a flow-induced mass term, which is proportional to vocal fold acceleration and represented by matrix Q_2. All three matrices are functions of the

jet velocity U_j, which was used as the primary control parameter. The jet velocity U_j was related to the subglottal pressure P_s by

$$P_s = \frac{1}{2}\rho_f U_j^2 \left(1 - \frac{H_s^2}{H_{\text{in}}^2}\right), \tag{3}$$

where H_s and H_{in} are the glottal width at the flow separation location and the glottal inlet, respectively. The mass, stiffness, and structural damping matrices were symmetric matrices whereas the three generalized force matrices were in generally asymmetric (Zhang *et al.*, 2007). Assuming $q_0 = q_0 e^{st}$, Eq. (2) was solved for the eigenvalues s and eigenvectors q_0 as a function of the jet velocity. Phonation onset occurs at the jet velocity for which the growth rate (or real part) of one eigenvalue first becomes positive (Zhang *et al.*, 2007).

Zhang *et al.* (2007) showed that the primary mechanism of phonation onset was the eigenmode-synchronizing effect of the flow-induced stiffness. The flow-induced damping and flow-induced mass generally play a minor role in phonation onset, particularly at moderate to high structural damping. To better illustrate the effect of flow separation on phonation onset, results will be presented first for cases in which the flow-induced mass, flow-induced damping, and structural damping matrices were excluded from Eq. (2), yielding

$$M\ddot{q} + (K - Q_0)q = 0. \tag{4}$$

The results will be then discussed for the full system Eq. (2), in which all three components of the fluid pressure and structural damping are included.

Note that the mean state of the coupled system, around which the flow equations are linearized, is a function of the driven subglottal pressure or airflow. Therefore, for each given jet velocity, a nonlinear steady-state problem has to be solved to obtain the mean state of the coupled system, based on which the matrices in Eq. (2) are evaluated and the linear stability analysis is performed. In this study, as in Zhang *et al.* (2007), the mean state of the coupled system corresponding to each given jet velocity was not solved for. Instead, it was assumed that, for each given jet velocity, the statically deformed geometry of the vocal fold and the glottal channel remained the same as that given at rest, and the linear stability analysis was performed on this same vocal fold/glottal channel geometry.

III. RESULTS

For the results presented in the following, the following nondimensional values for the model parameters were used (see definition in Fig. 1):

$$D = 1.43, \quad g = 0.0714, \quad T_{\text{inf}} = 1, \quad T_{\text{sup}} = 0.2, \tag{5}$$

$$\rho_f = 0.0012, \quad \nu = 0.47,$$

where ν is the Poisson's ratio of the vocal fold material. Equation (5) corresponded to the following physical values (the first four are the scaling variables):

$$\bar{T} = 7 \text{ mm}, \quad \bar{\rho}_{vf} = 1030 \text{ kg/m}^3, \quad \bar{E} = 3 \text{ kPa},$$

$$\bar{U}_{\text{scaling}} = 1.7 \text{ m/s},$$

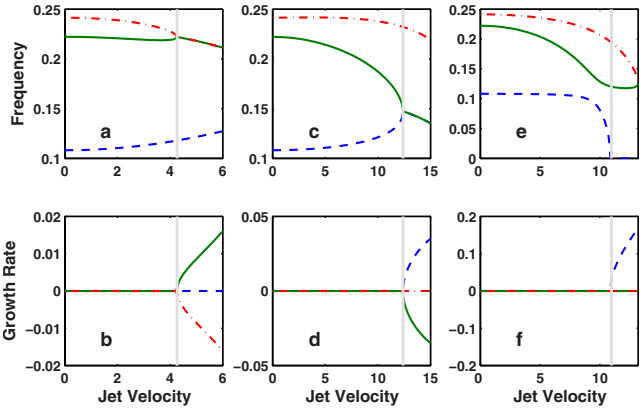


FIG. 2. (Color online) Frequencies and growth rates of the first three eigenvalues [First (---); second (—); and third (-·-)] of the coupled fluid–structure system [Eq. (4)] as a function of the jet velocity. The vertical line indicates the point of phonation onset. Straight glottal channel (α=0), model parameters specified in Eq. (5) were used. Only the flow stiffness term Q_0 of the three flow-induced terms was included [Eq. (4)], and σ=0. Flow separation location (a and b) at z=1.0 (superior edge of the medial surface); (c and d) z=0.2 (inferior part of the medial surface); and (e and f) z=1.2 (superior surface).

$$D = 10 \text{ mm}, \quad \bar{g} = 0.5 \text{ mm}, \quad \bar{T}_{\text{inf}} = 7 \text{ mm}, \tag{6}$$

$$\bar{T}_{\text{sup}} = 1.4 \text{ mm}, \quad \bar{\rho}_f = 1.2 \text{ kg/m}^3, \quad \nu = 0.47.$$

A. Coupled-mode flutter and eigenmode synchronization

Equation (4) may exhibit two types of instabilities: static divergence and coupled-mode flutter. For a conservative system, which corresponds to symmetric matrices M, K, and Q_0 in Eq. (4), the system loses stability first to static divergence (zero frequency) instability as the jet velocity increases from zero. For a nonconservative system for which the flow-induced stiffness matrix Q_0 is asymmetric (e.g., due to flow separation), the system may lose stability first to a coupled-mode flutter (nonzero frequency instability and therefore phonation onset), due to a cross-mode coupling effect of the asymmetric Q_0 matrix. An example is given in Figs. 2(a) and 2(b), which shows the frequencies and growth rates of the first three eigenvalues of Eq. (4) as the jet velocity increases, for a uniform glottal channel (divergence angle α=0) and a flow separation location at z=1.0 (superior edge of the medial surface). In this case, coupled-mode flutter occurred due to the synchronization of the second and third eigenmodes.

B. Effects of flow separation location

To illustrate the effects of flow separation location on phonation onset, the flow separation location was artificially manipulated along the vocal fold surface. A vocal fold with a parallel medial surface (α=0) was used, with model parameters given by Eq. (5). Figure 3 shows the phonation threshold jet velocity, threshold pressure, and onset frequency (circle symbols) as a function of the flow separation location. Also shown in Fig. 3 is the ratio between the glottal widths at the flow separation location and the minimum glottal constriction, or the separation constant as used in Lous *et al.* (1998).

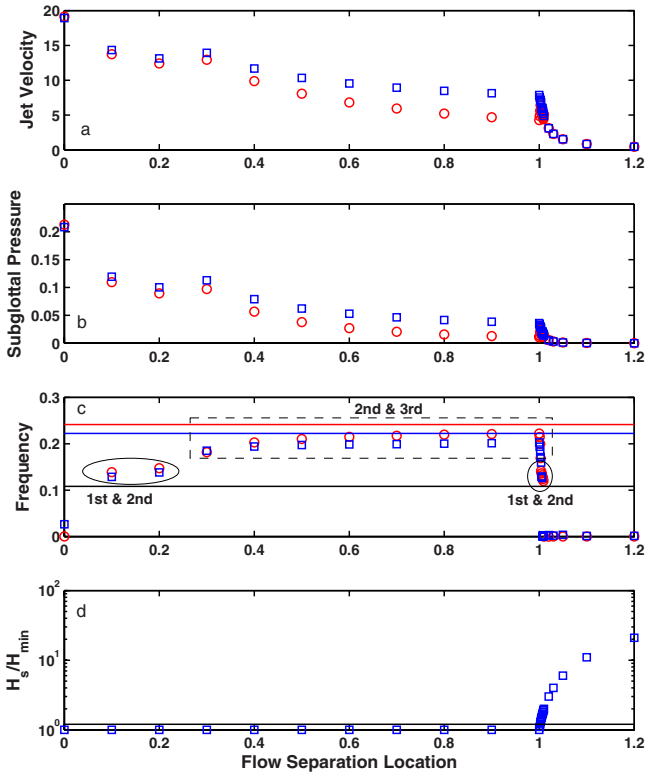


FIG. 3. (Color online) (a) Phonation threshold jet velocity, (b) threshold pressure, (c) onset frequency, and (d)glottal width ratio $H_s/H_{\min}$ as a function as the flow separation location, for a straight glottal channel (α=0). Model parameters are given in Eq. (5). Only the flow-induced stiffness Q_0 of the three flow components was included, and σ=0 (○) and all three flow components (Q_0, Q_1, and Q_2) were included, and σ=0.3 (□). The three horizontal lines in (c) indicate the first three *in vacuo* eigenfrequencies. The horizontal line in (d) indicates a constant glottal width ratio of 1.2.

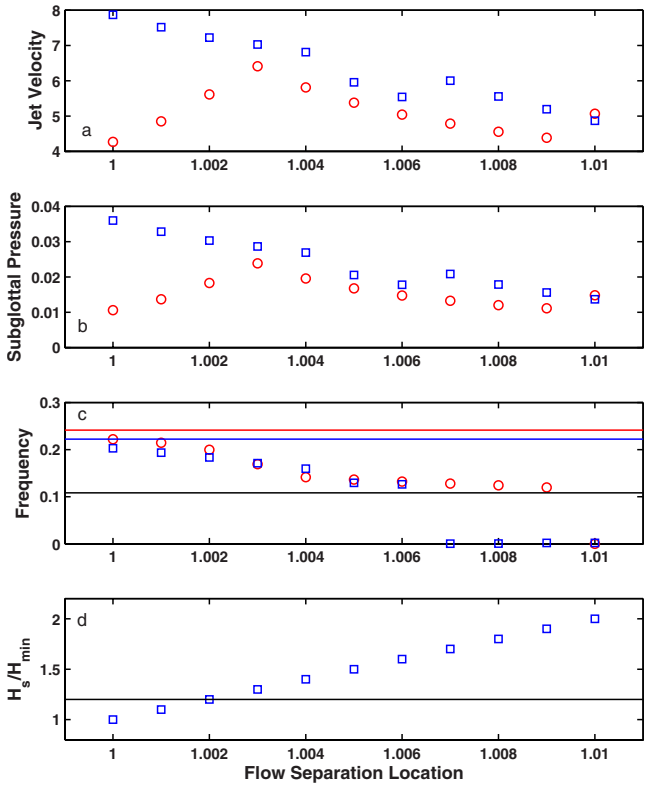


FIG. 4. (Color online) Close-up view of Fig. 3 for flow separation locations in the range of (1, 1.01).

Figure 3 shows that the flow separation location, which affects the intraglottal air pressure distribution, had a significant impact on the eigenmode-coupling effect of the flow-induced stiffness, which is the primary mechanism of phonation onset. As the fixed flow separation location varied along the vocal fold surface, the mode-coupling strength of the flow-induced stiffness also varied, leading to a different eigenmode-synchronization pattern. The system lost stability to different types of instabilities: both static divergence (z $>$1.01 and $z$$<$0.1, for a weak eigenmode-coupling effect) and coupled-mode flutter ($0<z<1.01$, for a strong eigenmode-coupling effect) were observed. For coupled-mode flutter instability, as the flow separation location was moved downstream, phonation onset may also occur due to eigenmode-synchronization involving different pairs of eigenmodes. For example, the second and third eigenmode synchronized for a flow separation location at z=1.0, as shown in Figs. 2(a) and 2(b). As the flow separation location moved either downstream or upstream along the vocal fold surface, the eigenmode-synchronizing pair changed to the first and second eigenmodes, as shown in Figs. 2(c) and 2(d) for a flow separation location at z=0.2. For very superior or inferior flow separation locations, the eigenmode-coupling effect was so weak that the system lost instability to a static divergence [Figs. 2(e) and 2(f)], indicating no phonation onset based on linear stability theory. For the case shown in Fig. 3, such change in the eigenmode-synchronization pattern occurred in two regions along the medial surface: One close to the inferior edge of the medial surface ($0<z<0.2$), and the other around the superior edge of the medial surface (around z=1.0). Note that the second transition region around z=1.0 is in a region where flow separation would be predicted to occur in many simplified models of flow separation.

Figure 4 shows the same results as in Fig. 3, but for a small range of flow separation locations around the superior edge of the medial surface (z=1.0), where flow separation would be predicted to occur in many simplified models of flow separation. In this region, eigenmode synchronization was highly sensitive to changes in the flow separation location. If a separation constant was used to determine the flow separation location, a slightly different separation constant in the range of (1, 1.9) would lead to quite a different phonation threshold pressure and phonation frequency. For even larger separation constants, phonation would not be even possible.

For synchronizations involving the same pair of eigenmodes, the vocal fold vibration pattern gradually varied with the flow separation location, but generally remained qualitatively similar. However, significant changes in the vocal fold vibration pattern occurred when eigenmode synchronization changed to a different pair of eigenmodes. For example, when the flow separation location moved from z=0.2 to z =0.3, the vibration pattern changed from a more in-phase medial-lateral motion to a more out-of-phase medial-lateral motion.

Figures 3 and 4 also show the phonation threshold jet velocity, threshold pressure, and onset frequency (square symbols) for Eq. (2), in which all three components (flow-

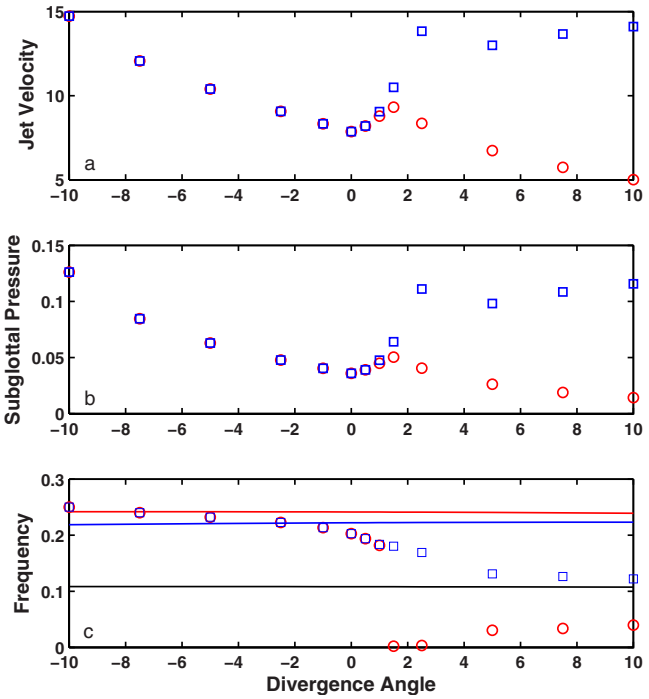


FIG. 5. (Color online) (a) Phonation threshold jet velocity, (b) threshold pressure, and (c) onset frequency as a function of the divergence angle. Negative values of the divergence angle indicate convergent glottal channels. Model parameters are given in Eq. (5). All three flow components (Q_0, Q_1, and Q_2) were included, and $\sigma=0.3$. Flow separation location fixed at the superior edge of the medial surface ($z=1.0$) (○) and flow separation location dependent on the divergence angle (□). The three horizontal lines indicate the first three *in vacuo* eigenfrequencies.

induced stiffness, damping, and inertia) of the flow pressure were included, and with a constant structural loss factor of 0.3. The result was similar to that when only the flow-induced stiffness and no structural damping were included. It is noted that, for a lower structural damping, the flow-induced damping introduced a destabilizing effect, which eventually destabilized the coupled system before the emergence of static divergence for flow separation locations ($z>1.01$) in which static divergence was predicted in Fig. 3. For larger values of the structural loss factor (as in Fig. 3), this destabilizing effect was delayed and occur at a higher jet velocity than the static divergence, and the coupled system still lost linear stability to static divergence.

C. Effects of glottal channel geometry

For a given vocal fold geometry, movement of flow separation location as discussed in the previous section is highly unrealistic. However, flow separation may occur at different locations for different vocal fold geometry. During one cycle of the vocal fold oscillation, the glottal channel alternately changes from convergent, straight, and divergent. Accordingly, the flow separation location may oscillate along the vocal fold surface. In particular, the flow separation location may move abruptly upstream when the glottis takes a divergent shape. In the following, the influence of flow separation within a divergent glottis is discussed.

Figure 5 shows the phonation threshold jet velocity, threshold pressure, and frequency as a function of the divergence angle of the glottal channel, as predicted from solving Eq. (2). Model parameter values are given in Eq. (5), and $\sigma=0.3$. Note that the vocal fold depth at the midpoint of the medial surface was kept constant to be $D=1.43$ as the divergence angle was varied. This helped to maintain a nearly constant *in vacuo* eigenspectrum of the vocal fold, thereby minimizing the influence of vocal fold natural frequencies on phonation threshold jet velocity. Similarly, the minimum static glottal half-width was also kept constant for different vocal fold divergence angles, excluding the effect of the glottal opening on phonation threshold jet velocity. The flow separation location was determined as follows: For convergent, straight, or slightly divergent glottal channels, flow separation was assumed to occur at the superior edge of the medial surface ($z=1.0$); for glottal channels of large divergence angle ($\tan\alpha>0.2g$), the flow separation location was assumed to be a location downstream of the minimum glottal constriction with a glottal width $H=1.2H_{\min}$, where $H_{\min}$ is the glottal width at minimum constriction.

Figure 5 shows that the minimum phonation threshold jet velocity (square symbols) occurred for nearly straight glottal channel geometry. In this case, the phonation threshold jet velocity increased slightly with increasing convergence of the glottal channel, whereas it increased rapidly when the glottal channel changed from convergent to divergent. Further increase in the glottal channel divergence angle led to only slight increase in the phonation threshold jet velocity. For convergent glottal channels, phonation onset occurred due to the synchronization of the second and third eigenmodes, whereas for large divergence angles, phonation occurred due to the synchronization of the first and the second eigenmodes. Consequently, the phonation threshold frequency decreased monotonically when the glottal channel changed from convergent to divergent, changing from being close to the second in vacuo eigenfrequency to being close to the first in vacuo eigenfrequency. For even large divergence angles (e.g., $a>15°$, not shown in Fig. 5), the eigenmode-coupling effect of the flow-induced stiffness was so weak that the system lost linear stability to static divergence instability.

Also shown in Fig. 5 are the phonation threshold pressure and frequency for the same vocal fold geometries, but for flow separation fixed at the superior edge of the medial surface ($z=1.0$), independent of the vocal fold geometry. For convergent glottal channels, as expected, the results were the same as in cases in which the flow separation location was allowed to vary with vocal fold geometry. However, for large divergence angles ($\alpha>1.5$), with flow separation fixed at $z=1.0$, the linear stability theory predicted static divergence and no phonation onset was possible. This contrasted with the coupled-mode flutter instability as predicted when the flow separation was allowed to move upstream. For even larger divergence angles, a low-frequency oscillation was induced by the flow-induced damping. However, the corresponding frequency was much lower than that of the coupled-model flutter when the flow separation was allowed to move upstream.

IV. DISCUSSION

The analysis was repeated for a few different vocal fold geometries. Although some details differed from those in Figs. 3 and 5 depending on the exact vocal fold geometry, structural damping, and glottal half-width used, the general features remained qualitatively similar. For example, a minimum phonation threshold pressure existed for a near-rectangular glottal configuration. Variation of the flow separation location along the vocal fold surface affected the eigenmode synchronization, which may enhance or reduce coupling in general or selectively affect coupling between certain modes. The mode-coupling effect was generally weaker for a divergent glottal configuration.

The results shown in Fig. 5 are consistent with previous experimental results in Chan *et al.* (1997). Their experimentally measured phonation threshold pressure showed a similar trend with varying glottal divergence angle (Fig. 4 in Chan *et al.*, 1997) as shown in Fig. 5 of this study. They observed that the lowest phonation threshold pressure was obtained for a rectangular or near-rectangular prephonatory glottis, which is consistent with the observation of this study. The measured phonation threshold pressure in Chan *et al.* (1997) in most cases increased at a faster rate with increasing divergence angle than with increasing convergence angle, which was qualitatively reproduced in Fig. 5 of this study.

The existence of an optimal glottal configuration associated with the lowest phonation threshold pressure was also predicted by Lucero (1998). However, some differences in the results between that study and the present study can be noted. First, Lucero showed that, for flow separation fixed at the superior edge [referred to as "no flow separation" in Lucero (1998)], no optimal glottal configuration existed and the phonation threshold pressure increased monotonically with decreasing divergence angle. However, this study showed that an optimal glottal channel configuration existed even when the flow separation was fixed at the superior edge, indicating its existence may be mostly due to a geometrical effect. Second, Lucero showed that allowing the flow separation location to move upstream for divergent glottal shapes increased the phonation threshold, prohibiting phonation. However, our results show that phonation was facilitated by this upstream movement of the flow separation location at large divergence angles, at which phonation otherwise would not be possible if the flow separation location was fixed at the superior edge of the medial surface. Further, our results clarified that, in this case, phonation was facilitated by a change in the eigenmode synchronization pattern as the glottal shape changed from slightly divergent to highly divergent. Note that at least three degrees-of-freedom are required to capture this change of synchronization pattern, which is not possible in a two degrees-of-freedom system. Future experimental investigations are needed to clarify these discrepancies.

V. CONCLUSIONS

In this study, we showed that, for a given glottal configuration, the phonation threshold pressure and frequency were highly sensitive to the location of flow separation. Although the actual movement of the flow separation point may be limited in real phonation, this high sensitivity to the flow separation location still points to the need for phonation models to be capable of accurate prediction of flow separation and the flow separation location. This is particularly important for modeling pathological phonation, in which flow may separate at different locations from the two vocal folds. Due to the large vocal fold-airflow density ratio, the vocal fold motion may have a large impact on the glottal flow field and the flow separation location (Sciamarella and Le Quere, 2008). Therefore, such phonation models may require a better and realistic representation of the vocal fold biomechanics and the glottal fluid–structure interaction.

ACKNOWLEDGMENTS

The author would like to thank Dr. Juergen Neubauer for many discussions and comments on the manuscript. This study was supported by research Grant Nos. R01 DC009229 and R01 DC003072 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

Alipour, F., and Scherer, R. C. (**2004**). "Flow separation in a computational oscillating vocal fold model," J. Acoust. Soc. Am. **116**, 1710–1719.

Chan, R. W., and Titze, I. R. (**1999**). "Viscoelastic shear properties of human vocal fold mucosa: measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021.

Chan, R. W., Titze, I. R., and Titze, M. R. (**1997**). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.

Ishizaka, K. (**1981**). "Equivalent lumped-mass models of vocal fold vibration," in *Vocal Fold Physiology*, edited by K. N. Steven and M. Hirano (University of Tokyo, Tokyo), pp. 231–244.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1267.

Kucinschi, B. R., Scherer, R. C., DeWitt, K. J., and Ng, T. T. M. (**2006**). "Flow visualization and acoustic consequences of the air moving through a static model of the human larynx," J. Biomech. Eng. **128**, 380–390.

Lous, N. J. M., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," Acust. Acta Acust. **84**, 1135–1150.

Lucero, J. C. (**1998**). "Optimal glottal configuration for ease of phonation," J. Voice **12**, 151–158.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Scherer, R. C., Shinwari, D., DeWitt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Sciamarella, D., and Le Quere, P. (**2008**). "Solving for unsteady airflow in a glottal model with immersed moving boundaries," Eur. J. Mech. B/Fluids **27**, 42–53.

Zhang, Z., Mongeau, L., Frankel, S. H., Thomson, S. L., and Park, J. B. (**2004**). "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**, 1720–1728.

Zhang, Z., Neubauer, J., and Berry, D. A. (**2007**). "Physical mechanisms of phonation onset: a linear stability analysis of an aeroelastic continuum model of phonation," J. Acoust. Soc. Am. **122**(4), 2279–2295.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, Part I: Computational methods and sound generation mechanisms," J. Acoust. Soc. Am. **112**, 2134–2146.

Perceptual context effects of speech and nonspeech sounds: The role of auditory categories

Radhika Aravamudhan[a)]
School of Audiology, Pennsylvania College of Optometry, Elkins Park, Pennsylvania 19020

Andrew J. Lotto
Speech, Language, and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

John W. Hawks
School of Speech Pathology and Audiology, Kent State University, Kent, Ohio 44242

(Received 27 July 2007; revised 22 April 2008; accepted 13 June 2008)

Williams [(1986). "Role of dynamic information in the perception of coarticulated vowels," Ph.D. thesis, University of Connecticut, Standford, CT] demonstrated that nonspeech contexts had no influence on pitch judgments of nonspeech targets, whereas context effects were obtained when instructed to perceive the sounds as speech. On the other hand, Holt *et al.* [(2000). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722] showed that nonspeech contexts were sufficient to elicit context effects in speech targets. The current study was to test a hypothesis that could explain the varying effectiveness of nonspeech contexts: Context effects are obtained only when there are well-established perceptual categories for the target stimuli. Experiment 1 examined context effects in speech and nonspeech signals using four series of stimuli: steady-state vowels that perceptually spanned from /ʊ/-/ɪ/ in isolation and in the context of /w/ (with no steady-state portion) and two nonspeech sine-wave series that mimicked the acoustics of the speech series. In agreement with previous work context effects were obtained for speech contexts and targets but not for nonspeech analogs. Experiment 2 tested predictions of the hypothesis by testing for nonspeech context effects after the listeners had been trained to categorize the sounds. Following training, context-dependent categorization was obtained for nonspeech stimuli in the training group. These results are presented within a general perceptual-cognitive framework for speech perception research. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956482]

PACS number(s): 43.71.An, 43.71.Es, 43.66.Ba, 43.66.Ki [PEI] Pages: 1695–1703

I. INTRODUCTION

The perceived phonemic category of a speech sound can be influenced by the surrounding phonetic context. One of the earliest demonstrations of the context-dependent nature of speech perception was by Lindblom and Studdert-Kennedy (1967). They presented Swedish listeners with three series of synthetic vowels. The first series consisted of steady-state vowels varying acoustically in the frequency of the second formant (F2) and varying perceptually from /ʊ/ (low F2) to /ɪ/ (high F2). The second and third series embedded the vowels in a time varying formant context that corresponded perceptually to /wVw/ and /jVj/ (where V stands for a vowel from the series). The resulting vowel categorization functions differed depending on the context condition. Relative to the steady-state condition, more /ɪ/ responses were obtained for the /wVw/ context and more /ʊ/ responses were obtained for the /jVj/ context. Similar context-dependent vowel categorization has subsequently been demonstrated by Nearey (1989) using stop consonant contexts (/bVb/ vs /dVd/) and by Nabelek and Ovchinnikov (1997), using isolated vowels versus /wV/ contexts (with no steady-state portion).

The perceptual context shifts obtained for vowels provocatively mirrors an effect of context witnessed in normal speech production. Lindblom (1963) reported that the acoustic realization of vowels is influenced by the surrounding consonants being produced. Because vowel articulation tends to assimilate toward the articulation of the context consonants, the formant frequencies "undershoot" the values that would be obtained when the vowel is produced in isolation. Thus, an /ɪ/ produced in the /wVw/ context results in a lower F2 than if produced in isolation (closer to the low F2 at the onset and offset of /w/). Thus a /wVw/ context in production results in acoustics (lower F2) that are more similar to /ʊ/ produced in isolation. Conversely, Lindblom and Studdert-Kennedy (1967) demonstrated that /wVw/ context results in more /ɪ/ perceptual categorizations. That is, /wVw/ context results in more /ʊ/-like *productions* but more /ɪ/-like *perceptions*. The context-dependent perception appears to compensate for context-dependent changes inherent in production.

The symmetry of speech perception and production suggests that the mechanisms underlying the perceptual context effects may be specific to speech sound processing. Lindblom and Studdert-Kennedy (1967) proposed that their context effect demonstrated "that *categorization of the continuum is adjusted* in the different *environments so as to compensate for an undershoot effect* in the vowel stimuli" (pp. 830 and 842; original italics). Repp (1982) suggested

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

that phonetic context effects are the result of listeners' "implicit knowledge of this coarticulatory variation (p. 97)." This perceptual compensation or implicit knowledge could be the result of experience with the acoustic effects of undershoot (coarticulation) or could be part of a specific perceptual process that has evolved to handle the complexity of speech production—acoustic mapping [e.g., a phonetic module such as proposed in later descriptions of motor theory (Liberman and Mattingly, 1989)].

An alternative explanation for the perceptual context effects was described by Repp (1982): Context effects result from general-auditory mechanisms (not specific to speech) with "the perception of relevant acoustic cues…somehow affected by the context" (p. 27). Repp (1982) rejected this account because "no plausible mechanisms that create such effects have been suggested, and no similar effects with nonspeech analogs have been reported so far" (p. 97). Subsequently, such an auditory account was tested by Holt *et al.* (2000) using /bVb/ and /dVd/ contexts similar to those utilized by Nearey (1989). Holt *et al.* (2000) first replicated the results of Nearey (1989) and Lindblom and Studdert-Kennedy (1967) by obtaining shifts in the categorization of a vowel series varying from /uh/ (as in but) to /eh/ (as in bet) as a function of consonant context (more /uh/ responses in /dVd/ context). They then replaced the formant transitions corresponding to the consonants with a single sine-wave tone varying in frequency along the path of F2 from the original speech stimuli. The rationale for this manipulation was to test whether context effects required the context to have phonetic content. If the perceptual effects of context are a result of a process that compensates specifically for coarticulation then shifts in categorization should only be obtained when the context is speech that can be coarticulated with the vowel. On the other hand, if perceptual shifts occur with contexts that only share some acoustic similarity with the speech contexts (but no phonetic content) then it indicates that general-auditory interactions may play a role. In fact, Holt *et al.* (2000) obtained a shift in categorization functions for the nonspeech contexts that was similar in size and direction as was obtained with the analogous speech contexts. In a follow-up experiment, Holt *et al.* (2000) also obtained perceptual shifts when the contexts were steady-state sine-wave tones with a frequency matched to the F2 onset frequencies of the transitions for the speech contexts.[1]

The pattern of results reported by Holt *et al.* (2000) is in line with experiments by Lotto *et al.* (1997) and Holt (2005) demonstrating nonspeech contexts affecting speech perception. Holt and Lotto (2006) have proposed that these context effects arise from general perceptual mechanisms that enhance the spectral contrast between neighboring sounds. A redescription of the perceptual results presented above makes this contrastive nature more apparent. In each of the cases, the vowel is categorized more often as the endpoint with the higher F2 frequency (/eh/ or /ɪ/) when the context has a lower-frequency F2 onset (/wVw/, /wV/, /bVb/, or sine waves modeling /bVb/ F2 transitions). Conversely, more low-frequency F2 vowel categorizations (ʊ/ or /uh/) occur in the contexts with high-frequency F2 onsets (/jVj/, /dVd/, or sine waves modeling /dVd/ F2 transitions). That is, the perception of the "relevant acoustic cue" (F2 frequency in the case of these vowel contrasts) is affected contrastively by the surrounding context whether it is speech or nonspeech. These contrastive perceptual patterns have also been demonstrated for speech contexts and nonspeech targets (Stephens and Holt, 2003) and for speech contexts and targets with human infant (Fowler *et al.*, 1990) and avian (Lotto *et al.*, 1997) listeners.

Whereas the results of Holt *et al.* (2000) appear to indicate that it is not critical that the context has phonemic content in order to induce speech categorization shifts, it is still the case that the listeners' task was phonetic identification. Therefore, any putative phonetic module would be active during the task making it difficult to determine if the perceptual effects are really nonspecific to speech perception. Williams (1986) conducted a study that attempted to disentangle phonetic and spectral contents in context effects by manipulating the listeners' task as opposed to the acoustics of the stimuli. Based on the stimulus sets of Lindblom and Studdert-Kennedy (1967), Williams (1986) constructed two sets of stimuli consisting of sine waves with frequencies that tracked the trajectories of the first three formants for /wʊw/-/wɪw/ and for /ʊ-ɪ/. That is, there was a /w/ context sine-wave series and an isolated-vowel sine-wave series with steps varying in the frequency of the middle tone corresponding to F2 in each series. For one task, Williams (1986) asked listeners to categorize these stimuli as exemplars of the vowels /ʊ/ and /ɪ/, taking advantage of the fact that listeners can perceive sine-wave replicas of speech phonetically when instructed to do so (e.g., Bailey *et al.*, 1997; Remez *et al.*, 1981; Best *et al.*, 1981). The categorization results demonstrated a similar context effect to that obtained by Lindblom and Studdert-Kennedy (1967). More /ɪ/ categorizations were obtained from the series modeling /wVw/ than from the series modeling isolated vowels. This result would be predicted by both speech-specific and general-auditory accounts of phonetic context effects. In the second task, listeners were presented the same stimulus sets but were asked to categorize the stimuli in terms of pitch: "high" versus "low." This judgment would presumably be based on the perceived frequency of the second tone (mimicking F2) because this is what varied in the series. If the perception of the frequency of F2 is altered by context, as proposed by Holt and Lotto (2006), then one should obtain a shift in pitch judgments toward more high judgments in the /wVw/ series as opposed to the isolate vowel series (a contrastive shift from the lower frequency of the context tones). In fact, Williams (1986) obtained no shift in relative pitch judgments when comparing context and isolated conditions. These results suggest that context effects require processes specific to phonetic perception. Nearey (1989) suggested that the Williams result (1986) provides strong evidence that these effects are the result of a "speech mode" as opposed to a general-auditorv effect.

The results of Williams (1985) and Holt *et al.* (2000) provide some problems for coherent interpretation. If context effects are specifically the result of phonetic perception, then why would nonspeech contexts have an effect on vowel categorization? One possibility is that the phonetic perceptual processes (or phonetic module) are not narrowly tuned and

nonspeech information can be integrated into phonetic judgments. However, this would not explain why birds show similar context effects with, presumably, no perceptual processes specifically tuned to human speech (Lotto *et al.*, 1997). Another problem with interpreting the Williams results (1986) is that Mullenix *et al.* (1988) in an attempted replication of Williams results (1986) did obtain more high pitch judgments for sine-wave stimuli based on /wVw/ compared to a series based on isolated vowels. Nevertheless, if context effects were the result of general perceptual processes, why would it be so much easier to obtain shifts in phonetic categorization tasks as opposed to more general tasks such as relative pitch judgments? One possible explanation is that the difference between these two tasks extends beyond the fact that one uses phonetic labels and the other pitch labels.

One of the salient differences between speech perception and other perceptual tasks studied by psychoacousticians is that adult listeners have extensive experience with speech sounds and that they have fairly well-defined perceptual categories for these sounds. Assigning high or low pitch labels to sounds is not a very typical task and, in our experience doing nonspeech studies, is not one that naive participants perform particularly well with no practice. In addition, it is unlikely that listeners have any well-established perceptual categories for novel sine-wave stimuli such as those used by Williams (1986).[2] It is possible that the lack of experience with the labeling task and existing pitch categories for the stimuli interfered with the ability to obtain context-dependent shifts in the pitch judgment task. With no established referent categories for pitch, the listeners may have adjusted their criteria for the labeling task to the perceptual range presented by the stimulus sets. That is, even if a general-auditory contrast effect was present for the context stimuli, the criterion for pitch judgments could be set relative to the shifted perceptual range because the criterion is not anchored to any predetermined category structure. The experiments presented below were designed to test the hypothesis that context effects can be obtained for nonphonetic categorization tasks if the listener has established perceptual categories specific to the labeling task. The comparisons made here are between two speech series and two nonspeech series (sine-wave analogs) that are modeled on acoustic characteristics of the speech series. In keeping with Lindblom and Studdert-Kennedy (1967), Williams (1986), and Mullenix *et al.* (1988), the context effects are measured as differences in response to isolated-vowel stimuli (/*u*/–/ɪ/) versus vowels in the context of /w/. However, in the current study, the context condition corresponds to /wV/ as opposed to /wVw/. These stimuli are modeled in part after parameters used by Nabelek and Ovchinnikov (1997) in their extensive study of the acoustic parameters that affect the size of context effect shifts. They demonstrated that preceding the vowel with /w/ results in significant shifts in vowel categorization but following the vowel with /w/ does not.

In experiment 1, participants were asked to label the nonspeech series with arbitrary labels (group 1 or group 2) with minimal practice and then were asked to label the speech series with the phonetic category for the vowel. Based on the results of Williams (1986), it was predicted that categorization shifts between context and isolated series only would be witnessed for the phonetic task. Experiment 2 provides participants with the same stimuli and tasks except that the listeners from one group go through extensive training categorizing the isolated-vowel sine-wave stimuli as groups 1 and 2 (with no experience with the context sine-wave stimuli). The prediction derived from the hypothesis presented above is that subsequent to training listeners will demonstrate a shift in categorization when the sine-wave stimuli are presented in /wV/ modeled context. This would indicate that the null effect for pitch judgments in the work of Williams (1986) may be due to the lack of well-formed perceptual categories for this task, as opposed to the nonphonetic nature of the task. However, if a context effect is not obtained for the nonspeech series in experiment 2 then this will provide further evidence [from better matched tasks than used by Williams (1986)] that context effects are specific to phonetic categorization.

II. EXPERIMENT 1

Experiment 1 is designed to investigate whether context effects can be elicited with a nonphonetic task. Context effects will be operationally defined as shifts in the overall percentage of a particular category response between an isolated (steady-state) series and a series with context. Categorization responses were collected for four series (two speech and two nonspeech sine-wave analogs). The four series were (1) isolated vowels synthesized to vary from /ʊ/ to /ɪ/ (V series), (2) vowels synthesized in the /wV/ context (wV series), (3) sine-wave analogs with steady-state frequencies at the nominal formant frequencies for the isolated vowels (V-SW series; vowel-sine wave), and (4) sine-wave analogs of the wV speech series (wV-SW series). Based on previous work by Nabelek and Ovchinnikov (1997) and Williams (1986), it is predicted that the phonetic categorization of the speech series will result in a context effect, but that the categorization of the nonspeech series will elicit no context-moderated shift.

A. Method

1. Subject

Twenty native English speaking adults, within the age range of 20–40 years, participated in return for course credit at Kent State University. None of the listeners were phonetically trained. All the subjects were screened for normal hearing (thresholds <20 dB Hearing Level (dBHL), 250–8 kHz).

2. Stimuli

a. Speech series Isolated vowel (V) and context (wV) series were constructed, each with 150 steps of variation in nominal F2 frequency. All speech stimuli were synthesized using a digital formant synthesizer in cascade mode (Klatt, 1980) with 16 bit resolution and a sampling rate of 10 kHz. The V series consisted of 200 ms steady-state formants and a rising fundamental frequency contour from 100 to 120 Hz. The frequency and bandwidth of F1 were fixed at 400 and 60 Hz, respectively, for all tokens. The frequencies of F4 and

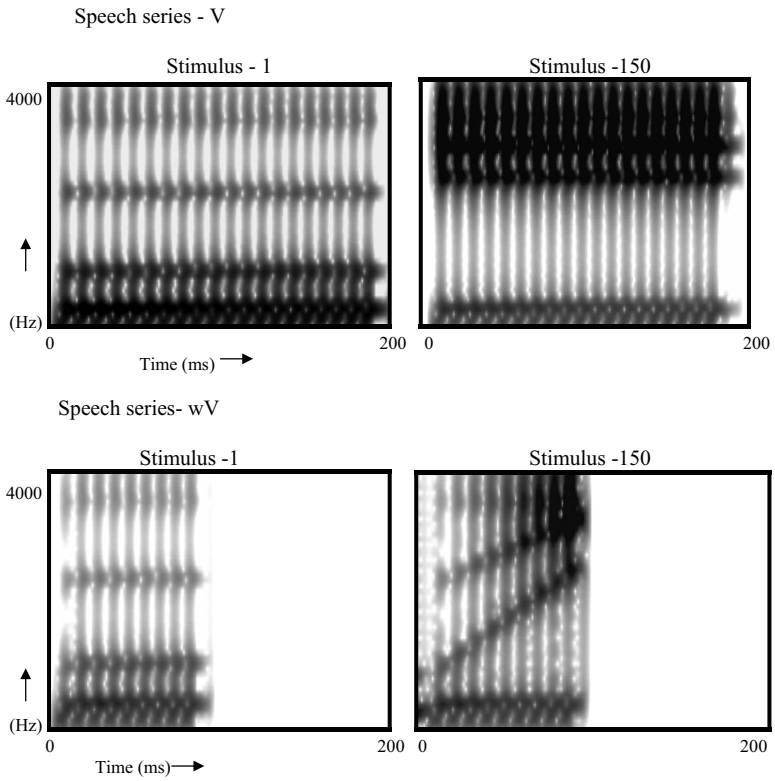


FIG. 1. Spectrograms showing endpoint stimuli for the speech series with isolated vowel (V; top) and vowel in /w/ context (wV; bottom).

F5 were fixed at 3300 and 3800 Hz with bandwidths of 250 and 300 Hz, respectively. The steps of the series varied in the frequency of F2 and F3. The frequency of F2 increased in 10 Hz steps from 1000 to 2500 Hz. These endpoint values matched the stimuli used by Nabelek and Ovchinnikov (1997). Lindblom and Studdert-Kennedy (1967) and Williams (1986) varied their F3 frequency as a function of F2. For the present stimuli, F3 frequency was calculated from F2 frequency using rule described by Nearey (1989).[3] The series varied perceptually from /ʊ/ to /ɪ/. The stimuli for the wV series started with a 20 ms steady-state portion for the formants appropriate for an initial /w/ with F1=250 Hz, F2 =800 Hz, and F3=2200 Hz and bandwidths of 60, 100, and 110 Hz, respectively (Nabelek and Ovchinnikov, 1997). The remaining 75 ms of the stimulus consisted of formant transitions from the initial /w/ values to the values described for the V series above. Thus, there was no steady-state portion for the vowels, only a transition to the final vowel target. This comparison of isolated steady-state stimuli with context stimuli lacking a steady-state portion mimics the characteristics of the stimuli set used by Lindblom and Studdert-Kennedy (1967). Nabelek and Ovchinnikov (1997) in a series of parametric studies determined that the frequency difference between initial and final F2 was the main determinant of shifts in category responses. Spectrograms of representative V and wV series stimuli are presented in Fig. 1.

b. Nonspeech series Sine-wave analogs mimicking the speech stimuli described above were constructed using the multispeech program from Kay-Elemetrics (model 2700, version 2.3). For the V-SW series, 200 ms steady-state sine waves were created with frequencies that matched the frequencies for F1, F2, and F3 of each V series token. For the wV-SW series, the sine waves varied in frequency to match the frequency patterns of the first three formants from each exemplar of the wV series. The intensity of each band was the same as that of the corresponding speech formant. For both series, the sine waves were added together with random starting phase using CoolEdit Pro (version 1.2, Syntrillium Corp.). The rms amplitude of each stimulus varied by less than ±0.6 dB.

3. Procedure

All testing was conducted in a sound-isolated booth with stimuli presented binaurally through headphones (Sennheiser HD25SP) at a comfortable listening level [~70 db sound pressure level (A)]. Stimulus presentation and response collection were controlled by a computer. Responses were made by mouse click on labeled boxes presented on a monitor. The response time was self-paced and participants could choose to listen to each stimulus as many times as they wanted before responding. The experiment consisted of four blocks of 150 stimuli. The order of the blocks was fixed for all participants, whereas the order of stimuli within a block was randomized. Each block contained stimuli from one of the four stimulus series. The order of presentation was V-SW, wV-SW (two nonspeech series), V and wV (two speech series). The nonspeech series were always presented first in order to dissuade listeners from perceiving the sounds as analogs of speech. They were told that they would hear tones that should be labeled as members of "group 1" and "group 2." No instructions were given as to the basis of this labeling difference. Prior to the testing on the V-SW series, participants were presented with a short practice set consisting of ten tokens of the endpoints from the V-SW series (five repetitions of each endpoint). Participants labeled these practice stimuli and were provided visual feedback ("correct" or "wrong"). Group 1 was mapped to the V-SW endpoint with the lowest second tone frequency (analog to F2), whereas group 2 was mapped to the stimulus with the highest second

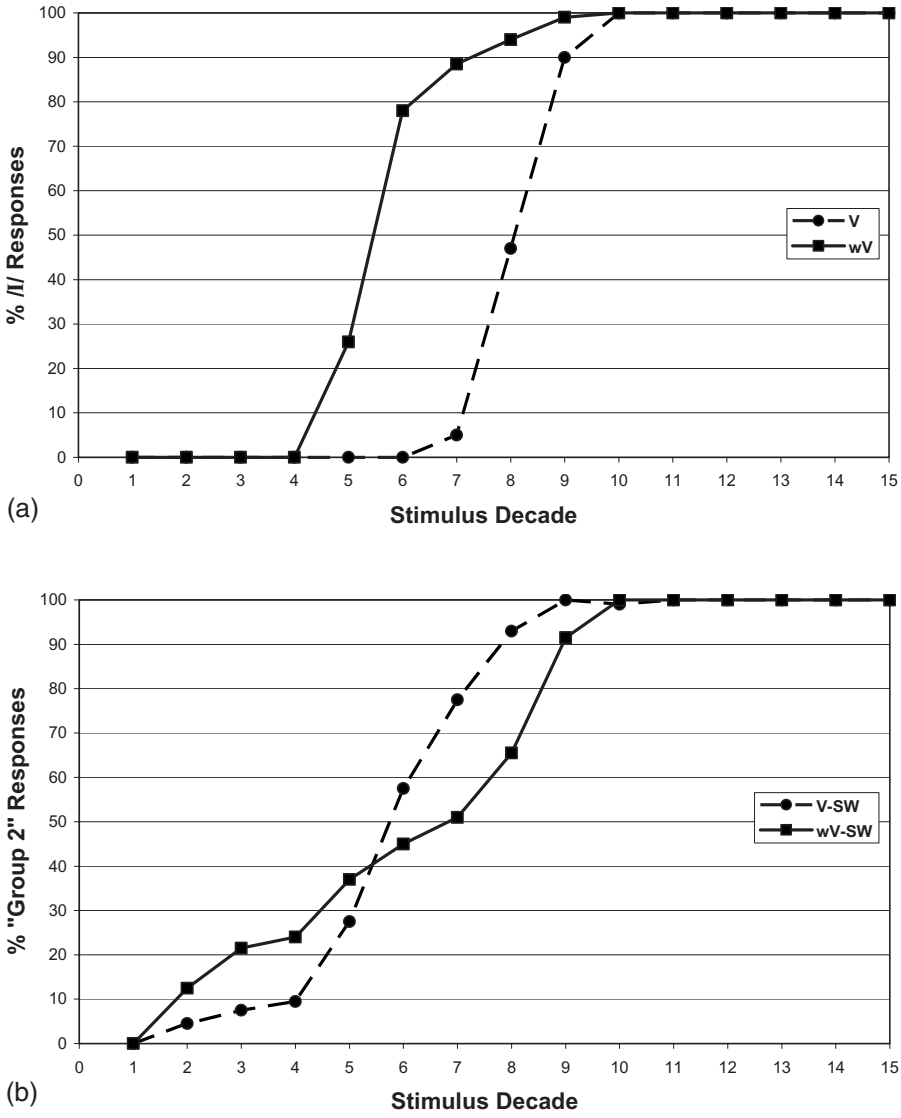


FIG. 2. Categorization functions for experiment 1. (a) Speech series: mean percentage of /ɪ/ responses by stimulus step decades (groups of ten stimuli: 1–10, 11–20, etc.) for isolated (V) and context (wV) conditions. (b) Nonspeech series: mean percentage of group 2 responses by stimulus decade for isolated (V-SW) and context (wV-SW) conditions.

tone frequency. This mapping corresponds to the low versus high pitch judgment utilized by Williams (1986) but avoids pre-existing labels. This short practice session was sufficient to help participants develop a basis for grouping of the nonspeech stimuli. No feedback was presented during any of the subsequent testing blocks. Following the V-SW block, participants labeled the wV-SW stimuli using the same group 1 and group 2 labels. The speech series blocks were then presented with new instructions to label the sounds as containing the vowel in "hood" (/ʊ/) or "hid" (/ɪ/). A short break was provided between each block and the entire session lasted approximately 40 min.

B. Results

The dependent variable of interest is the percentage of stimuli receiving responses corresponding to high-frequency stimuli (i.e., /ɪ/ and group 2 for speech and nonspeech series, respectively) collapsed across the 150 stimuli of each series. A context effect would be defined as a significant difference in percent categorization of /ɪ/ or group 2 between isolated (V or V-SW) and context condition (wV or wV-SW). The predicted direction of this effect would be an increase in "high-frequency" responses in the context conditions (counteracting the effects of vowel undershoot or a spectral contrast with the lower-frequency formants or tones in the context).

The first prediction to be tested is that speech stimuli will elicit a context effect. A paired-sample t-test revealed a significant difference in percent /ɪ/ responses obtained in the V (49.5%) and wV (65.6%) series [$t(19)=12.34$, $p<0.001$]. Figure 2(a) provides a graphic demonstration of the categorization shift. Categorization functions were computed for each subject by averaging the percent /I/ responses across each decade of stimuli (i.e., 1–10, 11–20, etc.) These individual categorization functions were averaged together to create the functions in Fig. 1(a). One can see a clear shift in the phonetic boundary due to context. These results replicate previous demonstrations of /w/ context-induced shifts on vowel categorization (Nabelek and Ovchinnikov, 1997).

The second prediction, based on the results of Williams (1986), was that nonspeech analogs would fail to induce a context effect. This prediction was supported by nonsignificant differences between the V-SW and wV-SW conditions in % of group 2 responses [65.1% vs 63.4%, $t(19)=2.01$, $p=0.06$]. (Note that the trend toward a significant effect for %

group 2 responses is actually in the opposite direction predicted for a context effect with more group 2 responses in the isolated condition.) This null effect is generally in line with the results of Williams (1986) with a slightly different stimulus set and a different task (arbitrary categorization as opposed to pitch judgments). However, the inability of nonspeech contexts to elicit categorization shifts in nonspeech targets is difficult to reconcile with demonstrations of nonspeech contexts influencing speech target categorization, such as those of Holt *et al.* (2000). Experiment 2 was designed to examine whether training on the nonspeech categorization task, with the presumed development of stable representations for these categories, will result in context-moderated categorization for nonspeech stimuli.

III. EXPERIMENT 2

In experiment 2, the same participants from experiment 1 were divided into "control" and "training" groups. The training group received additional exposure and feedback on the nonspeech categorization task. (In experiment 1, training consisted of only ten trials, five from each series endpoint). After training, the same categorization tasks were performed as in experiment 1 with the same dependent variables. The data from experiment 1 served as a comparison for any observed shifts subsequent to training in experiment 2. The hypothesis being examined here is the following: Well-formed target categories are necessary in order to obtain context effects. The first prediction resulting from this hypothesis is that, with no subsequent training, the control group will continue to fail to display context-moderated shifts in categorization performance for nonspeech stimuli. The second prediction is that the training group will show significant nonspeech context effects following training (whereas they showed no context effect in experiment 1).

A. Method

1. Subjects

The participants from experiment 1 were randomly assigned to the training and control groups. These groups did not differ on percent /I/ or percent high responses or boundary locations for any of the conditions in experiment 1 (tested with t-tests on each condition and dependent variable separately).

2. Stimuli

The stimulus set for the training was the V-SW (only steady-state sine-wave analogs were used) series from experiment 1. The post-training tasks included all four stimulus sets from experiment 1.

3. Procedure

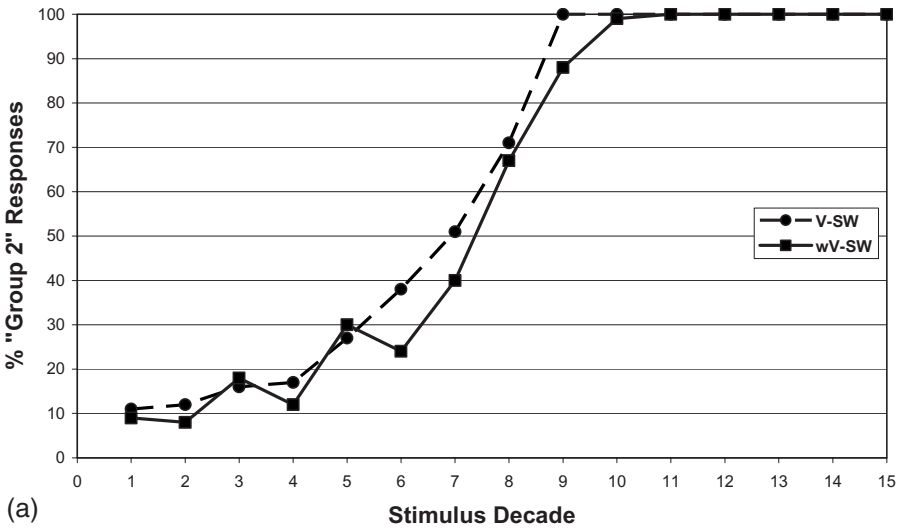
Members of the training group returned on a separate day from experiment 1 to begin training on the nonspeech categorization task. These participants were only trained on the V-SW stimuli. That is, they only received experience and feedback with the steady-state isolated sine-wave stimuli. A training session consisted of a single block during which each of the 150 stimuli was presented once in random order. All stimulus presentation details were equivalent to the description for experiment 1. After each stimulus presentation, the participant categorized the sound as a member of group 1 or group 2 by using a mouse to click a labeled response box on the computer monitor. After the response, feedback was provided visually (correct or wrong). The feedback was determined from a criterion that was specified separately for each participant as equal to their categorization boundary from experiment 1. These boundaries were defined as the stimulus number at which there was a change in response to at least three consecutive responses of the alternative category (remembering that each stimulus received only one categorization response from each participant). Correct responses were group 1 responses to stimulus steps less than their calculated boundary stimulus from experiment 1.

Members of the training group participated in three training sessions on consecutive days. Immediately following the third training session, a categorization test was performed. The test consisted of 30 stimuli randomly drawn from the 150 stimuli of the V-SW series. Participants categorized these stimuli with no feedback. If their performance was 90% correct or greater (correct relative to their experiment 1 boundary) then they were done with the training. If performance was lower than 90% then the participants returned for a subsequent training session. Testing followed training for every session starting at session 3. Training continued until 90% correct performance was obtained on the test or six training sessions were completed. Six of the ten subjects were able to obtain 90% correct identification in four sessions, while the four remaining subjects required five sessions to obtain 90% correct identification.

When training was completed for each participant in the training group they returned the following day (after training) to complete the categorization tasks for all speech and nonspeech sessions. Participants of the control group were asked to return the next day from the day they participated in their first set of categorization task to complete the categorization tasks for all (isolated and context) speech and nonspeech series. This session was identical to the experiment 1 session.

B. Results

The first prediction to be tested was that there would be no context effects for the nonspeech stimuli for the control group. These participants received no categorization training and it was expected that they would categorize the nonspeech sounds similarly to how they had in experiment 1 (no effect of context). There was a small but significant context-moderated difference in % group 2 responses [63.7% vs 60%, $t(9)=2.83$, $p<0.05$], but the shift was in the opposite direction obtained in normal context effects, as was the insignificant shift in experiment 1. The categorization functions for the no-training group are presented in Fig. 3(a). These results support the expectation that the control group would show a little change in categorization performance from experiment 1 to experiment 2.



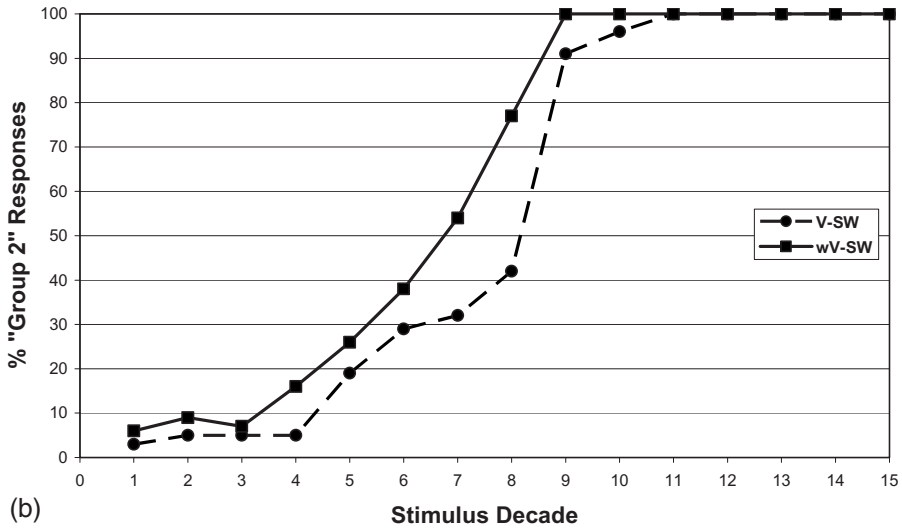


FIG. 3. Nonspeech series categorization functions for experiment 2. Mean percentage of group 2 responses are presented by stimulus decade for isolated (V-SW) and context (wV-SW) conditions. (a) Nontraining group; (b) training group.

Of greater importance for the hypothesis is the prediction that after training with nonspeech categorization, members of the training group would provide significant context shifts for the nonspeech stimuli. This prediction was supported by a significant context effect for % group 2 responses [55.1% vs 62.2%, $t(9)=4.88$, $p<0.001$]. The categorization functions for the training group are presented in Fig. 3(b). These results demonstrate that nonspeech contexts can shift the perception of nonspeech targets when there is adequate experience with the categorization task.

No explicit predictions were made regarding the speech series following training, but it was presumed that speech context effects would be maintained for both the training and control group. The shift in % /I/ as a function of context /w/ was maintained for both groups. [V: 48% vs wV: 66.6%, $t(9)=7.73$, $p<0.000$ (training group) and V:47.9% versus wV: 63.7%, $t(9)=7.67$, $p<0.000$ (control group)].

IV. DISCUSSION

The goal of the current study was to test a hypothesis that may reconcile the seemingly contradictory results of Williams (1986) and Holt *et al.* (2000). Williams (1986) demonstrated that nonspeech contexts had no influence on pitch judgments of nonspeech targets, but that context effects were obtained when the listeners were instructed to perceive the sounds as speech. On the other hand, Holt *et al.* (2000) reported that nonspeech contexts with no phonemic content could shift the categorization of speech targets. Why would nonspeech contexts be capable of influencing the perception of speech sounds but not the perception of similar nonspeech sounds? One hypothesis is that the well-developed categories for the target speech sounds (vowel phonemes) allowed the effects of context to become manifest, whereas the novel nonspeech stimuli and pitch judgment task of Williams (1986) did not have the necessary well-established perceptual categories. The two experiments described in this paper were designed to test predictions arising from this hypothesis.

Experiment 1 once again demonstrated that speech sound categorization is influenced by context, whereas novel nonspeech categorization tasks are not. Listeners categorized vowels (/ɪ/ versus /ʊ/) presented as isolated steady states and in the context of a preceding /w/. As demonstrated previously by Nabelek and Ovchinnikov (1997), the percentage of /ɪ/ responses increases in /w/ context. In contrast, the categorization of sine-wave complexes as group 1 and group 2 did

not differ between isolated and context (sine-wave transitions analogous to the formant transitions of /w/) conditions.

Experiment 2 examined whether training on the nonspeech categorization task would result in context-sensitive perception of the nonspeech targets. Participants were trained to categorize the isolated nonspeech targets based on a criterion (placed at their individual categorization boundaries from experiment 1) or received no subsequent training or exposure to the stimuli. The control group continued to show disparate patterns of categorization for speech and nonspeech stimuli. The training group, on the other hand, demonstrated context-dependent shifts in nonspeech categorization that were in the same direction as the speech context effects.

The results of experiment 2 demonstrate that context effects can be obtained for nonspeech stimuli. Holt *et al.* (2000) demonstrated the effects of nonspeech contexts on the categorization of speech sounds. Stephens and Holt (2003) obtained effects of speech contexts (/al/ and /ar/) on the discrimination of nonspeech sounds (distinguishing F2-F3 combinations filtered from /da/-/ga/ stimuli). The present study adds to this literature by exhibiting an effect of nonspeech context on nonspeech categorization. All of these context effects can be considered contrastive in the manner described by Holt and Lotto (2006). In the case of the current study, the frequency of the second tone, which models F2 in the speech stimuli, was the dimension of importance for the categorization of the isolated steady-state stimuli. The context sound consisted of tones increasing in frequency to the values of the steady-state tones. That is, the context included energy at a lower frequency than the steady-state second tone. The result according to Holt and Lotto (2006) would be that the critical tone would be perceived as effectively higher in frequency. This would lead to more group 2 (trained to correspond to the stimuli with higher second tone frequencies) responses in the context condition, which is exactly what was obtained.

The fact that these contrastive context effects were obtained with nonspeech contexts and targets are in line with the proposition that contrast is the result of the operating characteristics of the general auditory system (Holt and Lotto, 2002; Lotto *et al.*, 2003). If contrast is a consequence of the way that the auditory system encodes complex sounds then it likely plays some role in context effects, such as those presented by Lindblom and Studdert-Kennedy (1967), Nearey (1989), and Nabelek and Ovchinnikov (1997). A general auditory basis for this effect makes it less surprising that phonetic context effects occur in infants (Fowler *et al.*, 1990), have been demonstrated in birds (Lotto *et al.*, 1997), and appear to be independent of the native language of the listener (Mann, 1986; Gow and Im, 2004).

One of the pieces of evidence against the general perceptual nature of these effects was the lack of a context effect for the nonspeech pitch judgments in Williams, 1986. The results of the current study suggest that the failure to find context effects was not because the sounds were perceived as nonspeech, but because there were not well-established perceptual categories on which to base the pitch judgments. When listeners are trained to categorize the stimuli based on a fixed criterion (as in experiment 2), nonspeech context can affect nonspeech categorization as predicted by a spectral contrast account. This raises the question of what aspect of the training was responsible for the manifestation of context-sensitive categorization. We speculate that when listeners are asked to do a relative judgment task such as high or low pitch or group 1 versus group 2 with no established categories, they place their decision boundary to around the middle point of the (perceptual) stimulus range. Such a default strategy has been witnessed in other studies on the categorization of nonspeech sounds based on center frequency of noises (Sullivan *et al.*, 2005a, 2005b). Because this criterion is a function of the auditory (as opposed to acoustic) range of stimuli, it is free to shift with changes in the perceived range. In the context condition, according to the spectral contrast account, the effective pitch values of the critical second tones are increased relative to the lower-frequency context. If the entire range is affected then the midpoint of that range in terms of the acoustic value (stimulus step) is unchanged. The result is no shift in judgments due to context.

On the other hand, if one receives training in the form of feedback relative to a fixed acoustic boundary, the result is that the boundary is fixed to the corresponding auditory value, independent of the entire range. Sullivan *et al.* (2005a, 2005b) have shown that listeners can learn to shift their categorization criterion based on feedback rather quickly for novel nonspeech sounds. In the subsequent context condition, the shift in effective second tone pitch results in more stimuli being perceived as larger than the fixed boundary value. The result is more group 2 judgments and a significant context effect.

The results presented here are another indication of the fruitfulness of examining the roles of both the general-auditory and learning/cognitive systems when attempting to explain the perception of complex sounds such as speech (Holt *et al.*, 2004; Diehl *et al.*,. 2004). Lindblom (1996) had discussed the Williams (1986) study as demonstrating that perception is not "merely" about responding to an auditory signal, but is a function of "signal-plus-knowledge." That is, our interpretation of an acoustic stimulus is always a function of our peripheral representation of that signal and the application of a knowledge base relative to that signal. The present results support this point even more strongly than the original Williams (1986) study. The differences in perception of speech and nonspeech are not inherent in the nature of the signal; they are the result of our experience with and of the functional classification of these signals.

[1]This result allayed concerns that the effect was the result of auditory "capture" of some harmonics by the context tone (Dannerbring, 1976; Darwin, 1984; Darwin and Sutherland 1984).

[2]We have referred to the Williams (1986) tasks as phonetic categorization and pitch judgment to make the point about the differences in these two tasks.

[3]Nearey's (1989) rule for F3 calculation: For $F2 > 1500$, $F3_{front} = 0.522 \times F1 + 1.197 \times F2 + 57$; $F2 <= 1500$, $F3_{back} = 0.7866 \times F1 - 0.365 \times F2 + 2341$.

Bailey, P. J., Summerfield, A. Q., and Dorman, M. (**1977**). "On the identification of sine wave analogues of certain speech sounds," Haskins Laboratory Status Report on Speech Research, Report No. SR-51052, Yale University, New Haven, CT.

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," Percept. Psychophys. **29**, 191–211.
Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.
Darwin, C. J., and Sutherland, N. S. (**1984**). "Grouping frequency components of vowels: When is a harmonic not a harmonic?," Q. J. Exp. Psychol. A **36**, 193–208.
Dannerbring, G. L. (**1976**). "Percieved auditory continuity with alternately rising and falling frequency transitions," Can. J. Psychol. **30**, 99–115.
Diehl, R. L., Lotto, A. J., and Holt, L. L. (**2004**). "Speech perception," Annu. Rev. Psychol. **55**, 149–179.
Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). "Young infants' perception of liquid coarticulatory influences on following stop consonants," Percept. Psychophys. **48**, 559–570.
Gow, D. W., Jr., and Im, A. M. (**2004**). "A cross-linguistic examination of assimilation context effects," J. Mem. Lang. **51**, 279–296.
Holt, L. L. (**2005**). "Temporally non-adjacent non-linguistic sounds affect speech categorization," Psychol. Sci. **16**, 305–312.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.
Holt, L. L., and Lotto, A. J. (**2002**). "Behavioral examinations of the neural mechanisms of speech context effects," Hear. Res. **167**, 156–169.
Holt, L. L., Lotto, A. J., and Diehl, R. L. (**2004**). "Auditory discontinuities interact with categorization: Implications for speech perception," J. Acoust. Soc. Am. **116**, 1763–1773.
Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–990.
Liberman, A. M., and Mattingly, I. G. (**1989**). "A specialization for speech perception," Science **243**, 489–494.
Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.
Lindblom, B. E. F., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.
Lindblom, B. (**1996**). "Role of articulation in speech perception: clues from production," J. Acoust. Soc. Am. **99**, 1683–1692.
Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)," J. Acoust. Soc. Am. **102**, 1134–1140.
Lotto, A. J., Sullivan, S. C., and Holt, L. L. (**2003**). "Central locus of non-speech context effects on phonetic identification," J. Acoust. Soc. Am. **113**, 53–56.
Mann, V. A. (**1986**). "Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r"," Cognition **24**, 169–196.
Mullennix, J. W., Pisoni, D. B., and Goldinger, S. D. (**1988**). "Some effects of time-varying context on the perception of speech and nonspeech sounds," Report No. 14, Indiana University, Bloomington, IN.
Nabelek, A. K., and Ovchinnikov, A. (**1997**). "Perception of nonlinear and linear formant trajectories," J. Acoust. Soc. Am. **101**, 488–497.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.
Stephens, J. D. W., and Holt, L. L. (**2003**). "Preceding phonetic context affects perception of nonspeech," J. Acoust. Soc. Am. **114**, 3036–3039.
Sullivan, S. C., Lotto, A. J., Newlin, E. T., and Diehl, R. L. (**2005a**). "Sensitivity to stimulus distribution characteristics in auditory categorization," J. Acoust. Soc. Am. **117**, 2596.
Sullivan, S. C., Lotto, A. J., Newlin, E. T., and Diehl, R. L. (**2005b**). "Sensitivity to changing stimulus distribution characteristics in auditory categorization," J. Acoust. Soc. Am. **118**, 1896.
Williams, D. R. (**1986**). "Role of dynamic information in the perception of coarticulated vowels," Ph.D. thesis, University of Connecticut, Stanford, CT.

Effects of spectral modulation filtering on vowel identification[a)]

Chang Liu[b)]
Department of Communication Sciences and Disorders, University of Texas at Austin, 1 University Station A1100, Austin, Texas 78712

David A. Eddins[c)]
Department of Otolaryngology, University of Rochester, 601 Elmwood Avenue, Box 609, Rochester, New York 14642 and International Center for Hearing and Speech Research, 52 Lomb Memorial Drive, Lyndon Baines Johnson Building, Room 2620, Rochester, New York 14623

(Received 13 September 2007; revised 4 June 2008; accepted 9 June 2008)

The goal of this study was to measure the effects of global spectral manipulations on vowel identification by progressively high-pass filtering vowel stimuli in the spectral modulation domain. Twelve American-English vowels, naturally spoken by a female talker, were subjected to varied degrees of high-pass filtering in the spectral modulation domain, with cutoff frequencies of 0.0, 0.5, 1.0, 1.5, and 2.0 cycles/octave. Identification performance for vowels presented at 70 dB sound pressure level with and without spectral modulation filtering was measured for five normal-hearing listeners. Results indicated that vowel identification performance was progressively degraded as the spectral modulation cutoff frequency increased. Degradation of vowel identification was greater for back vowels than for front or central vowels. Detailed acoustic analyses indicated that spectral modulation filtering resulted in a more crowded vowel space (F1 × F2), reduced spectral contrast, and reduced spectral tilt relative to the original unfiltered vowels. Changes in the global spectral features produced by spectral modulation filtering were associated with substantial reduction in vowel identification. The results indicated that the spectral cues critical for vowel identification were represented by spectral modulation frequencies below 2 cycles/octave. These results are considered in terms of the interactions among spectral shape perception, spectral smearing, and speech perception. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956468]

PACS number(s): 43.71.Es, 43.71.An, 43.66.Ba [MSS] Pages: 1704–1715

I. INTRODUCTION

Many acoustic features influence vowel identification, including absolute and relative formant frequency, formant transition, overall duration, fundamental frequency, spectral contrast, and spectral tilt. Some investigators have suggested that vowel formants exclusively provided the essential acoustic information for vowel identity (e.g., Hillenbrand *et al.*, 1995; Hillenbrand and Nearey, 1999), while other investigators have emphasized the importance of overall spectral shape on vowel identification (e.g., Bladon and Lindblom, 1981; Bladon, 1982; Zahorian and Jagharghi, 1993). The purpose of the present study is to investigate the influence of changes in global spectral shape on speech perception by systematically manipulating the spectral modulation content of vowel stimuli and evaluating the impact of such manipulations on vowel identification. Specifically, a goal was to determine if degradation of the vowel spectrum produced by parametric changes in the spectral modulation domain yields systematic changes in vowel identification. This investigation is the first in a series of studies designed to evaluate and optimize novel algorithms for manipulating speech in the spectral modulation domain.

Formant frequencies, especially F1 and F2, have been demonstrated to be primary acoustic cues for vowel perception (Miller, 1989; Nearey, 1989; Syndal and Gopal, 1986), although there is a high degree of variability for formant frequencies within each vowel category produced across talkers and phonetic contexts (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). Both the absolute formant frequency and the relative frequency of formant peaks, such as the distance between F2 and F1 and distance between F3 and F2, provided reliable cues for vowel perception, especially for steady-state vowels. The relative frequencies of formant peaks, combined with their relative amplitudes, largely determine the gross spectral properties of the vowel. Using an automatic vowel classification algorithm for vowels embedded in a consonant-vowel-consonant (CVC) context, Zahorian and Jagharghi (1993) directly compared the relative importance of absolute formant frequency and global spectral shape in vowel classification. They reported that vowel classification based on global spectral features was superior to classification based on formants, independent of the availability of F0 information. Furthermore, corresponding perceptual measures of vowel identification indicated a stronger correlation between perceptual confusions and automatic classification based on the global spectral shape than classification based on formants alone. Similarly, Bladon and Lindblom (1981) evaluated an auditory model of vowel perception based on the assumption that vowel identification

[a)]Portions of the data were presented at the 147th meeting of the Acoustical Society of America (J. Acoust. Soc. Am. **115**, 2231).
[b)]Electronic mail: changliu@mail.utexas.edu. Author to whom correspondence should be addressed.
[c)]Electronic mail: david_eddins@urmsc.rochester.edu.

was dependent on a spectral representation of loudness density (phons/Bark) versus cochlear place (Bark). Each stimulus representation was treated as a single spectral shape and the auditory perceptual distance for vowel pairs was measured. The model predictions closely matched the perceptual judgments of the quality difference between vowel pairs, consistent with the hypothesis that overall spectral shape characterized the qualitative and perceptual distances associated with different vowel spectra.

Global spectral shape involves the absolute formant frequency and the relative separation among vowel formants as well as the relative amplitudes of the spectral peaks and their associated valleys (i.e., spectral contrast). Several studies have investigated the effect of spectral contrast on vowel perception (Leek *et al.*, 1987; Alcantara and Moore, 1995). Using vowel-like harmonic complex in which the intensity level of harmonics corresponding to formant frequencies was incremented adaptively, Leek *et al.* (1987) measured the minimum peak-to-valley contrast sufficient for vowel identification. Their normal-hearing listeners required about 3 dB peak-to-valley contrasts to achieve vowel identification accuracy above 90%. Similarly, Alcantara and Moore (1995) reported that the identification of vowel-like harmonic complexes by normal-hearing listeners was nearly 80% correct for peak-to-valley contrasts of 4 dB corresponding to F1–F3 when vowel sounds were presented at 65 dB sound pressure level (SPL).

Spectral tilt, or the relative amplitude of high- to low-frequency components, is another important acoustic characteristic of global spectral shape. Changes in spectral tilt lead to significant changes in vowel quality. Ito *et al.* (2001) measured the identification of isolated, synthetic, steady-state Japanese vowels in which the F2 peak was removed while the amplitude of the remaining formant peaks was preserved. Despite changes in vowel quality, complete removal of the F2 peak did not alter vowel identification, suggesting that a spectral peak corresponding to F2 was not required for the accurate identification of Japanese vowels. To separate the effects of global spectral tilt and formant frequency, Kiefte and Kluender (2005) manipulated those parameters independently using American English vowels. In addition to producing a much more crowded vowel space, their results showed that both local (formant frequency) and global (spectral tilt) acoustic cues had strong perceptual effects and that effect of changing spectral tilt was mitigated as the temporal pattern of formant frequencies was retained. They concluded that effects of spectral tilt on vowel identification might be very limited, given that naturally spoken English vowels were generally dynamic.

Rather than direct manipulation of acoustic features (e.g., formant frequency and formant amplitude), another approach to the study of spectral features involves modification of the global spectral shape of speech stimuli. Several investigations have manipulated the global spectrum via spectral envelope smearing in an effort to either simulate reduced spectral resolution associated with hearing impairment or to determine the relative importance of various acoustic features to speech perception. In previous studies, spectral smearing has been accomplished by processing speech stimuli through a series of bandpass filters in the audio-frequency domain and systematically varying the filter width and/or slope (e.g., van Veen and Houtgast, 1985; ter Keurs *et al.*, 1992; 1993; Baer and Moore, 1993; 1994; Moore *et al.*, 1997), by low-pass or high-pass filtering in the temporal modulation domain (e.g., Drullman *et al.*, 1994a, 1994b, 1996), or by varying the number of audio-frequency channels in the context of cochlear implant simulations (e.g., Turner *et al.*, 1999; Fu and Nogaki, 2005; Liu and Fu, 2007). Using each of these techniques, speech perception in quiet or in noise can be degraded as a result of spectral smearing. Interestingly, Drullman *et al.* (1996) demonstrated that both low-pass and high-pass filtering in the temporal modulation domain were analogous to a uniform reduction in the spectral modulation domain, at least over the spectral modulation range below 2 cycles/octave. van Veen and Houtgast (1985) argued that this spectral modulation frequency range was the most important for speech perception and Qian and Eddins (2008) showed that the most important spatial information carried by head-related transfer functions corresponded to spectral modulation frequencies below 2 cycles/octave.

The global spectral envelope can be represented by an array of spectral modulation frequency components via Fourier transformation of the magnitude spectrum (i.e., second-order Fourier transform). A vowel spectrum contains many spectral modulation components, while a series of peaks and valleys that vary sinusoidally across audio frequency is represented by a single component in the modulation spectrum. Amplification or attenuation of spectral modulation frequency components results in changes in the global spectral shape, and may give rise to corresponding perceptual changes. The work of Drullman *et al.* (1994a, 1994b, 1996) and van Veen and Houtgast (1985) illustrates the importance of spectral modulation content to speech intelligibility; however, the relative importance of different regions in the spectral modulation domain below 2 cycles/octave has not been established. Clearly spectral smearing results in alterations in the spectral modulation domain; however, the acoustic and perceptual consequences of the global changes resulting from direct manipulations in the spectral modulation domain on the local spectral features of speech stimuli, such as formant frequency, spectral contrast, and spectral tilt, have not been quantified.

The present study was designed to examine acoustic and perceptual impact of filtering low spectral modulation frequency on vowel sounds. A more complete understanding of the relationship between manipulations of the spectral modulation domain and speech perception will guide future studies involving simulations of various types of hearing impairment as well as studies of the efficacy of speech enhancement algorithms based on modifications in the spectral modulation domain. Accordingly, 12 American English vowels were synthesized and the spectral contrast, spectral tilt, and spectral details of these stimuli were manipulated simultaneously by progressively adjusting the cutoff of a high-pass filter applied in the spectral modulation domain. Identification of these vowel stimuli was then measured as a function of filter cutoff frequency to evaluate effects of attenuating low spectral modulation frequency components on

vowel perception. To gauge the spectral shape perception abilities of each of the listeners included in this study using nonspeech stimuli, spectral modulation detection thresholds were measured over a wide range of spectral modulation frequencies and the resulting spectral modulation transfer functions were compared to laboratory norms and published data using similar measures of spectral shape perception.

II. METHOD

A. Listeners

Five adult native speakers of American English between the ages of 23 and 40 years served as listeners and were paid for their participation in this study. All listeners had normal-hearing sensitivity corresponding to pure-tone thresholds ≤15 dB Hearing Level (HL; ANSI, 2004) at octave intervals between 250 and 8000 Hz.

B. Stimuli

All stimuli were generated using a digital array processor (TDT AP2) and a 16 bit D/A converter with a sampling period of either 81.92 μs (12 207.0125 Hz; vowel identification) or 24.4 μs (40 983 Hz; spectral modulation detection). Following D/A conversion, stimuli were low pass filtered (5000 Hz for vowels or 18 000 Hz for spectral modulation), attenuated (TDT PA4), and routed via headphone buffer (TDT HB6) to an earphone (Etymotic ER-2) inserted in the left ear of each listener. Stimulus generation and data collection were controlled by the SYKOFIZX® software application (TDT). All testing was conducted in a double-walled sound attenuating chamber (IAC).

1. Spectral modulation detection stimuli

Spectral envelope perception was evaluated by measuring spectral modulation detection thresholds (peak-to-valley contrast in decibels) as a function of spectral modulation frequency (0.25–8.0 cycles/octave) superimposed on a broadband-noise carrier (200–12 800 Hz) following the technique described by Eddins and Bero (2007). The complex spectrum of the desired signal was computed as follows. Two 8192-point buffers, X (real part of the spectrum) and Y (imaginary part of the spectrum), were filled with the same sinusoid computed on a logarithmic frequency scale with a starting phase randomly selected from a uniform distribution between 0 and 2π rad on each presentation. Following scaling of the sinusoid to the appropriate modulation depth, the X and Y buffers were multiplied by independent 8192-point samples from a Gaussian distribution. Then the X and Y buffers were mutiplied by an 8192-point buffer filled with values corresponding to the magnitude response of a second-order Butterworth filter with a passband from 200 to 12 800 Hz. An inverse fast Fourier transform (FFT) was then performed on the complex spectrum (X,Y) to produce the desired 400-ms spectrally-shaped noise waveform. Finally, the waveforms were shaped with a 10-ms $\cos^2$ window and scaled to the desired presentation level. This resulted in independent noise stimuli with the appropriate spectral shape on each observation interval. The spectrum level of the flat-spectrum standard stimuli was 35 dB SPL. The overall level of the spectrally modulated stimuli was adjusted to be equal to that of the flat-spectrum standard. As a result, the levels of the peaks in the modulated spectrum exceeded 35 dB SPL slightly, in a manner dependent on the required spectral modulation depth. The top panel of Fig. 1 illustrates the magnitude spectrum of a spectrally modulated noise as described above. In this example, the sinusoidal spectral modulation has a frequency of 1 cycle/octave and consists of six cycles spanning the six octaves from 200 to 12 800 Hz. In this example, the modulation depth (i.e., peak-to-valley difference in decibels) is 10 dB and the modulation phase (θ) is π rad with respect to the low-frequency edge of the noise carrier. In the lower panel, the second-order FFT illustrates spectral modulation depth as a function of spectral modulation frequency. In this example, there is a peak at 1 cycle/octave.

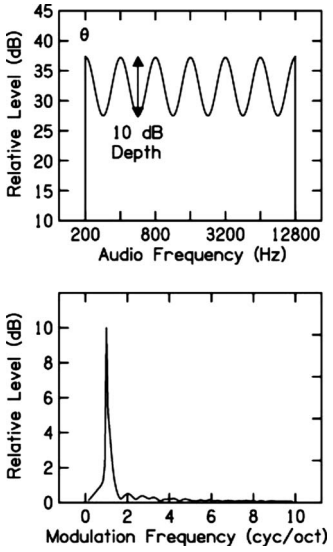


FIG. 1. Schematic representation of sinusoidal spectral modulation superimposed on a broadband-noise carrier from 200 to 12 800 Hz. In this example, the upper panel shows six modulation cycles over a span of six octaves, representing a spectral modulation frequency of 1 cycle/octave. The spectral modulation depth (peak-to-valley difference in decibels) is 10 dB and the modulation phase (θ) is π rad relative to the low-frequency edge of the noise carrier. The lower panel shows the corresponding modulation magnitude spectrum computed by taking a FFT of the magnitude spectrum shown in the upper panel.

2. Vowel identification stimuli

Vowel identification was measured for 12 American English vowels /i, ɪ, e, ɛ, æ, ʌ, ɜ̃, ɑ, ɔ, o, u, U/. Vowels were spoken by an adult, middle-aged female talker in the syllable context /bVd/ and were recorded and stored as digital sound files. Vowels were isolated by deleting the formant transition at the beginning and end of the syllable such that only the relatively steady-state vowel nucleus remained. These stimuli served as standard vowels. Vowel analysis and resynthesis were performed in MATLAB® as follows. First, the linear predictive coding (LPC) spectrum was computed from a 50 ms window of the central nucleus of the standard vowel using 16 LPC coefficients with a modified version of the Colea MATLAB® code (Loizou, 2000). Following conversion to a $\log_2$ frequency scale, the LPC spectrum was transformed

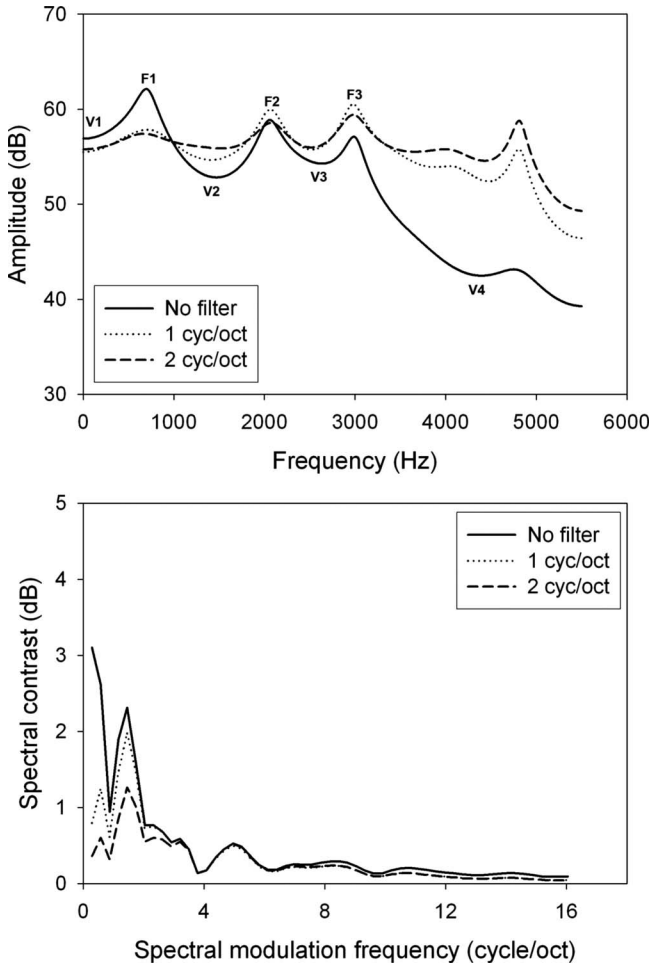


FIG. 2. LPC spectra (top) and spectral modulation spectra (bottom) for the unfiltered /æ/ vowel (solid) and /æ/ vowels filtered at 1 (dotted) and 2 cycles/octave (dashed).

to the spectral modulation domain via FFT. The resulting magnitude spectrum was high-pass filtered with cutoff frequencies of 0.0 (no filtering), 0.5, 1.0, 1.5, and 2.0 cycles/octave using a first-order digital Butterworth filter. Vowel spectra, with and without spectral modulation filtering, were reconstructed via inverse FFT of the modified magnitude spectrum and the preserved phase spectrum. This reconstructed spectrum was then duplicated to create a three-dimensional spectrogram with a total duration of 167 ms (the average duration of the 12 vowels spoken by the female talker). The final vowel waveforms were produced by resynthesis using a modified version of STRAIGHT (Kawahara *et al.*, 1999), in which the spectrogram of the standard vowel was replaced by the new spectrogram with the manipulation in the spectral modulation described above. The fundamental frequency of all vowel stimuli was normalized to 160 Hz (the average F0 of the 12 vowels for the female talker). This F0 was constant over the vowel duration.

All vowels were scaled to have equal root-mean-squared amplitudes. Each standard vowel had five derivatives (high-pass cutoffs of 0.0, 0.5, 1.0, 1.5, and 2.0 cycles/octave), yielding a total of 60 vowels (12 vowel categories $\times 5$ filter conditions). The sound pressure level of the vowels was set to 70 dB SPL. Figure 2 shows examples of unfiltered and high-pass filtered (1.0 and 2.0 cycles/octave) /æ/-vowel stimuli with the LPC spectrum shown in the upper panel and the magnitude spectrum in the spectral modulation domain shown in the lower panel. It is clear from the lower panel that the amplitudes of the low spectral modulation frequencies (<3 cycles/octave) are substantially higher than the amplitudes of the high spectral modulation frequencies. This is consistent with the report by van Veen and Houtgast (1985) that spectral smearing that approximated low-pass filtering in the spectral modulation domain above about 2 cycles/octave had little influence on vowel spectra or vowel perception.

C. Procedures

1. Spectral modulation detection

For the spectral modulation detection task, a three-interval, single-cue, two-alternative, forced-choice procedure was used to measure spectral modulation detection thresholds. On each trial, the three observation intervals were separated by 400-ms silent intervals. The standard stimulus was presented in first interval as an anchor or reminder. A second standard stimulus and the signal stimulus were randomly assigned to the two remaining presentation intervals. The threshold was estimated using an adaptive psychophysical procedure in which the spectral modulation depth was reduced after three consecutive correct responses and increased after a single incorrect response. The step size was initially 2 dB and was reduced to 0.4 dB after three reversals in the adaptive track, estimating 79.4% correct detection (Levitt, 1971). Spectral modulation detection thresholds were based on the average of three successive 60-trial runs.

2. Vowel identification

Vowel identification was measured in one unfiltered condition and four high-pass filtered conditions as described above. For a given condition, vowel identification performance was estimated on the basis of three 120-trial blocks with each of the 12 vowels presented ten times in a random order within a block. As a result, vowel identification performance in each condition was based on 360 trials consisting of 30 repetitions of each of the 12 vowel stimuli. Thus, percent correct vowel identification was based on 30 repetitions of each vowel stimulus for each listener. Short breaks were provided between blocks and all of the conditions were completed in four sessions with each session lasting from 1 to 1.5 h. Listeners were seated in front of a computer monitor that displayed the 12 response alternatives as a text box labeled with the English word corresponding to each vowel in a /hvd/ context (e.g., had, head, hayed, heed, hid, hod, hoard, hoed, hood, who'd, hud, and heard). Listeners responded by using a computer mouse to click on the button corresponding to their response choice. After each vowel presentation, the listener was allowed 10 s to respond. Prior to data collection, listeners were familiarized with the task over the course of a half-hour practice session using standard stimuli. During familiarization, feedback was provided to indicate the correct response on each trial. No feedback was provided during formal testing.

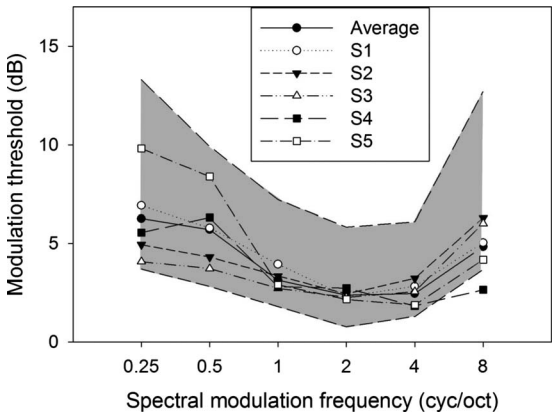


FIG. 3. Individual and average SMTFs for the five listeners in this study within the 95% confidence interval based on a group of 50 young, normal-hearing listeners (shaded region from unpublished data).

III. RESULTS

A. Spectral modulation detection

Spectral shape perception was evaluated by means of a spectral modulation detection task. Figure 3 shows spectral modulation detection threshold (peak-to-valley difference in decibels) as a function of spectral modulation frequency (cycles/octave) for each of the five listeners as well as the average spectral modulation transfer function (SMTF). In general, thresholds were lowest for the middle modulation frequencies and increased at very low and very high modulation frequencies. This pattern of results is consistent with the results of Eddins and Bero (2007). The shaded region displays the 95% confidence interval based on a group of 50 young, normal-hearing listeners who were inexperienced in psychoacoustic listening tasks and were unpracticed in the spectral modulation detection task (unpublished data). Most of the thresholds for the listeners of the current study fall within this range and thus may be characterized as having "typical" spectral envelope perception.

B. Vowel identification

Figure 4 shows the identification scores associated with each vowel category along the *x* axis for the vowels with and without high-pass filtering in the spectral modulation frequency domain (the upper panel) and the reduction in vowel identification due to the spectral modulation filtering (the lower panel), as indicated by the bars. For all the 12 vowels, identification scores decreased as the cutoff frequency of the high-pass spectral modulation filter increased up to 2 cycles/octave. Identification performance averaged across all vowel categories dropped dramatically as the high-pass filter cutoff increased from 0 cycle/octave (leftmost black bars: 92.9%) to 2 cycles/octave (rightmost shaded bars: 57.6%). The extent to which the identification performance decreased as a result of spectral modulation filtering was vowel specific (see the lower panel of Fig. 4). Attenuating low spectral modulation frequencies resulted in greater reductions in identification performance for the five back vowels (/ɑ, ɔ, o, u, U/) than the five front vowels (/i, ɪ, e, ɛ, æ/) or the two central vowels (/ʌ, ɜ̃/). As summarized in Fig. 5,

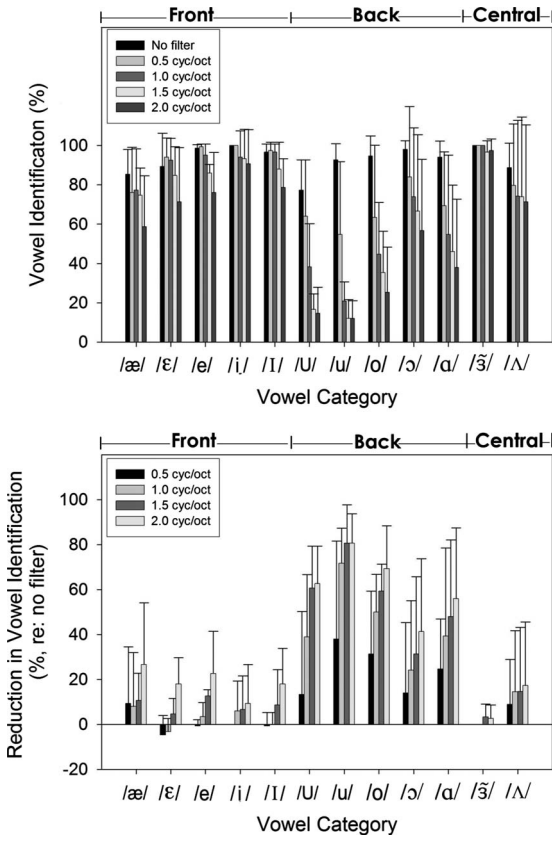


FIG. 4. Identification score of the vowels with and without spectral modulation filtering for all the 12 vowels (the upper panel) and reduction in identification score (regarding the unfiltered condition) of the four filtered conditions for the 12 vowel categories (the lower panel). The error bar stands for the standard deviation across the five listeners for each condition.

identification scores were quite similar for the front, back, and central vowels without filtering, while identification scores were markedly lower for the back vowels than the front or central vowels for the filtered conditions. Specifically, the three back vowels /o/, /U/, and /u/ showed substan-

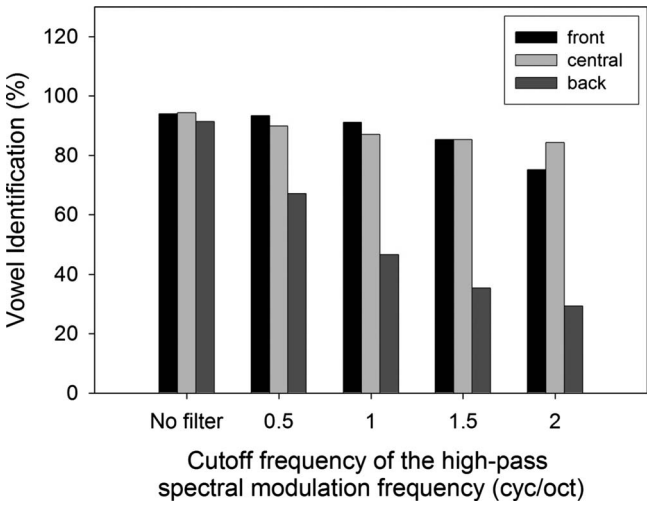


FIG. 5. Identification score as a function of the cutoff frequency of high-pass spectral modulation filter for three vowel groups: front, central, and back.

TABLE I. Confusion matrices for unfiltered vowels (top) and filtered vowels at 1 cycle/octave (middle) and 2 cycles/octave (bottom).

Vowel	æ	ɛ	e	i	ɪ	U	u	o	ɔ	a	ə	ʌ
æ	85.3%	14.0%									0.7%	
ɛ	8.0%	89.3%	0.7%	0.7%	1.3%							
e	0.7%		98.6%		0.7%							
i				100.0%								
ɪ		2.7%	0.7%		96.6%							
U						77.3%	4.7%	18.0%				
u						7.3%	92.7%					
o						4.7%		94.6%		0.7%		
ɔ									98.0%			2.0%
a									2.0%	94.0%		4.0%
ə											100.0%	
ʌ										11.3%		88.7%
æ	77.3%	22.7%										
ɛ	6.3%	92.5%		0.7%	0.5%							
e	0.5%	2.0%	95.2%		1.0%			1.3%				
i		1.3%	1.3%	94.0%	0.7%							2.7%
ɪ		2.0%	1.3%		96.7%							
U			12.5%	1.8%	0.5%	38.3%	5.2%	35.7%	0.7%	0.7%		4.6%
u			12.2%	5.2%	1.0%	31.5%	20.8%	22.7%	0.7%	2.7%		3.2%
o		0.7%	36.9%		3.0%	10.8%	1.3%	44.7%		1.3%		1.3%
ɔ	6.5%		1.5%			4.7%	0.7%	2.0%	73.8%	4.5%		6.3%
a	10.0%	0.5%	1.0%			0.7%			2.7%	54.7%		30.4%
ə											100.0%	
ʌ	9.5%	1.5%						0.5%		14.3%		74.2%
æ	58.7%	32.0%	0.7%		0.7%				3.3%	3.3%		1.3%
ɛ	11.3%	71.3%	0.7%	1.3%	0.7%					2.7%		12.0%
e	0.7%	1.3%	76.0%		13.3%	2.7%		4.0%		2.0%		
i		0.7%	2.0%	90.7%	1.3%				0.7%	2.0%		2.6%
ɪ	0.7%	16.6%	4.0%		78.7%							
U	0.7%		24.0	9.3%	1.3%	14.7%	4.0%	32.0%	2.6%	4.0%	0.7%	6.7%
u	1.3%	0.7%	24.7%	12.0%	3.3%	17.3%	12.0%	22.0%	0.7%	3.3%		2.7%
o	0.7%		52.0%	3.3%	7.3%	4.7%	0.7%	25.3%	2.7%	2.0%		1.3%
ɔ	8.0%	2.0%	4.0%	4.7%				5.3%	56.7%	9.3%	10.0%	
a	12.0%	1.3%	2.0%	2.6%		0.7%			4.7%	38.0%		38.7%
ə	2.0%					0.7%					97.3%	
ʌ	15.3%	0.7%	1.3%	0.7%						10.7%		71.3%

tial degradation in identification with spectral modulation filtering while the vowels /ɜ̃/ and /i/ showed little effect of spectral modulation filtering.

A nonparametric repeated measure analysis (Friedman's test) revealed that identification performance was significantly different across spectral modulation filtering conditions (chi square=113.465, $p<0.001$). Tukey multiple comparison tests indicated that vowel identification for the unfiltered vowels was significantly higher than any of the four filtered vowel conditions (all $p<0.01$). Separate nonparametric (Friedman's test) analyses were computed for front, central, and back vowels, indicating a significant effect of spectral modulation filtering on vowel identification for the five front vowels (chi square=44.701, $p<0.01$) and the five back vowels (chi square=72.958, $p<0.01$) but not for the two central vowels (chi square=5.419, $p=0.221$). In addition, Tukey multiple comparison tests indicated that for the front vowels, the unfiltered vowels were significantly more identifiable than vowels with filter cutoff frequencies of 1.5 and 2.0 cycles/octave ($p<0.05$) but not vowels filtered at 0.5 and 1.0 cycle/octave ($p>0.05$). For the back vowels, all four filtered cutoff frequencies led to significantly lower identification scores than the unfiltered vowels (all $p<0.05$).

Confusion matrices for the unfiltered, 1 cycle/octave, and 2 cycles/octave filter conditions are provided in Table I. Listeners showed few errors for the unfiltered vowels (upper section) while error rates increased and confusion was distributed more broadly across vowels as the high-pass cutoff frequency increased (middle and lower sections). In general, vowels were confused with their adjacent counterparts in the vowel space (see Fig. 6) and errors increased with filter cutoff frequency. For example, when the target vowel /æ/ was presented, listeners responded /ɛ/ incorrectly on 14% of the trials for the unfiltered condition. This error rate increased to 23% for the 1 cycle/octave condition and 32% for the 2 cycles/octave condition. As shown in Table II, more vowels were added to the confusion list for each vowel as the filter cutoff frequency increased. In addition, the vowel most frequently confused with the target vowel in the unfiltered condition remained as the most frequently confused vowel

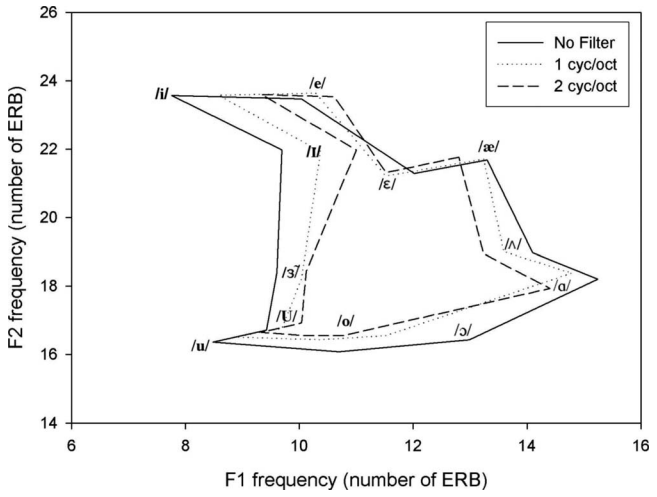


FIG. 6. Steady-state F1 and F2 frequencies in ERB scale for unfiltered vowels and vowels filtered at 1 and 2 cycles/octave.

for the filtered conditions. For example, identification of the intended vowel /U/ was primarily confused with /o/ in the unfiltered condition and was confused mainly with four vowels /e, i, o, ʌ/ for the 2 cycles/octave condition although the vowel /o/ remained the most frequently confused vowel. Moreover, it should be noted that the confusion matrices showed asymmetry between front and back vowels. For the filtered conditions (e.g., 2 cycles/octave), back vowels were frequently confused with front vowels (35.4%) while front vowels were rarely confused with back vowels (4.1%).

C. Correlation between SMTF, and vowel identification and its reduction

The error bars in Fig. 4 illustrate the variability across listeners in vowel identification and the associated reduction due to the spectral modulation filtering. Likewise, Fig. 3 reveals considerable variability in spectral modulation detection, particularly for low modulation frequencies. To examine the potential relationship between spectral modulation detection and vowel identification, correlation analyses were conducted between spectral modulation thresholds and vowel identification for the 12 individual vowel categories, the three vowel groups (front, back, and central), and for performance across all vowel groups. When collapsed across vowel category or vowel group, no significant correlations were obtained between spectral modulation thresholds and vowel identification (all $p>0.05$). Although these data do not indicate any relationship between spectral modulation detection and vowel identification associated with spectral modulation filtering, it is possible that the number of listeners in the present study was too small to reveal any underlying relationships.

D. Acoustic analyses

The progressive reduction in vowel identification scores with increasing high-pass spectral modulation cutoff frequency may result in simple changes in the acoustic signal (e.g., spectral contrast, tilt, and formant frequency shifts) or a series of acoustic changes that may or may not be vowel category dependent. This section will focus on the acoustic effects of spectral modulation filtering and the corresponding perceptual impact of those acoustic changes. The LPC spectra of all of the unfiltered and filtered vowels were obtained with the modified Colea MATLAB® code (Loizou, 2000) using 16 LPC coefficients over a 50 ms window of the vowel central nucleus. Formant frequency and amplitude, amplitude of the valleys adjacent to the first three formants (F1–F3), and spectral tilt were measured for all vowel spectra.

1. Formant frequency

Formant frequencies F1 and F2 for the unfiltered and filtered vowels were measured from the LPC vowel spectra. Because frequency is represented on a nonlinear scale by the auditory system, formant frequencies F1 and F2 were converted to an equivalent rectangular bandwidth (ERB) scale (Moore and Glasberg, 1987). A one-factor (filter) analysis of variance (ANOVA) with formant frequency (in ERB units) as the dependent variable indicated no significant change in formant frequency across filter conditions for either F1 ($F_{4,44}=0.361$, $p=0.835$) or F2 ($F_{4,44}=1.671$, $p=0.174$). As shown in Fig. 6, however, a nonsignificant difference in F1 may be due to the fact that F1 increased for vowels with low F1 frequency and decreased for vowels with high F1 frequency as the filter cutoff increased.[1] In this case, the posi-

TABLE II. Major confusion vowels (confusion percentage >5%) for unfiltered and filtered vowels. For multiple confusion vowels, the vowels that gave the highest confusion percentage were bold.

Vowel	No filter	0.5	1	1.5	2
æ	ɛ	ɛ	ɛ	ɛ	ɛ
ɛ	æ		æ	æ	æ, ʌ
e				ɪ	ɪ
i					
ɪ					ɛ
U	o	**u**, o	e, u, **o**	e, u, **o**, ʌ	e, i, **o**, ʌ
u		**U**, o	e, i, **U**, o	**e**, U, **o**	**e**, i, U, o
o		e, **U**	**e**, U	**e**, U	e
ɔ		ʌ	**æ**, ʌ	**æ**, ɑ, ʌ	æ, ɑ, ʌ
ɑ		**æ**, ʌ	æ, ʌ	æ, ʌ	æ, ʌ
ɝ̃					
Λ	ɑ	**æ**, ɑ	æ, **ɑ**	æ, **ɑ**	**æ**, ɑ

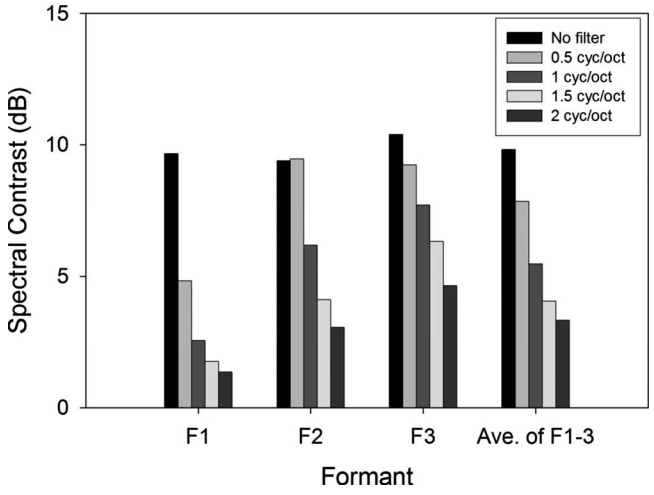


FIG. 7. Average spectral (peak-to-valley) contrasts in decibels over the 12 vowels for F1, F2, and F3, respectively, and for average of F1, F2, and F3 at the five cutoff frequencies of high-pass spectral modulation filter.

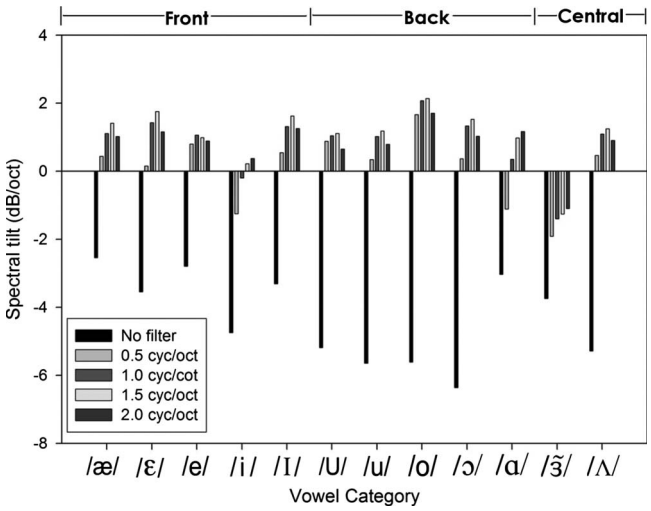


FIG. 8. Spectral tilt of the vowels at the five cutoff frequencies of high-pass spectral modulation filter for all the 12 vowels.

tive and negative changes canceled each other. The overall absolute change in formant frequency was 9.4% for F1 and was 1.4% for F2 for the 2 cycles/octave filter condition. A one-factor (filter) ANOVA based on the absolute percent change in formant frequency indicated significant change in formant frequency with spectral modulation filtering for both F1 ($F_{4,44}$=24.071, $p<0.05$) and F2 ($F_{4,44}$=6.202, $p<0.05$). *Post hoc* (Tukey) tests showed that all four filter conditions resulted in significant changes in formant frequency (all p values <0.05). The $F1\times F2$ vowel space also was reduced as the high-pass filter cutoff frequency increased from 1 to 2 cycles/octave (i.e., differences among F1 frequencies were reduced; see Fig. 6). Specifically, the back vowels were much closer in the vowel space for the 2 cycles/octave filter condition than for the unfiltered condition. Thus, it is possible that reduced vowel identification for back vowels could partially, if not all, result from increased crowding in the vowel space produced by high-pass filtering in the spectral modulation domain.

2. Spectral contrast for formants

As shown in Fig. 1, spectral (peak-to-valley) contrast, especially for F1 and F2, was progressively reduced as the attenuation of spectral modulation frequencies increased from low to high. The spectral contrast of vowel formants based on LPC spectra was computed as follows. The amplitudes corresponding to the valley below F1, between F1 and F2, between F2 and F3, and between F3 and F4 were defined as V1, V2, V3, and V4, respectively (see Fig. 2). The amplitudes corresponding to the spectral peaks were defined as F1–F3. Because each formant has two adjacent valleys, one below and one above the peak, the spectral contrast for F1–F3 was computed as the average of the two peak-to-valley contrasts (e.g., the average of F2-V2 and F2-V3 for F2), for all the unfiltered and filtered vowels. The overall spectral contrast for F1–F3 for the unfiltered and filtered conditions is shown in Fig. 7. Spectral contrast was substantially reduced for F1–F3. Specifically, as the filter cutoff frequency increased from 0.0 to 2.0 cycles/octave, the spectral contrast decreased from 9.7 to 1.4 dB for F1, from 9.4 to 3.0 dB for F2, and from 10.4 to 4.6 dB for F3. A one-way ANOVA indicated that spectral modulation filtering significantly reduced spectral contrast for all three formants (F1, $F_{4,44}$ =37.782, $p<0.05$; F2, $F_{4,44}$=10.339, $p<0.05$; F3, $F_{4,44}$ =11.631, $p<0.05$). Tukey *post hoc* tests revealed a significant difference in spectral contrast for F1 between the unfiltered vowels and all four filtered vowels (all p values <0.05), while the decrease in spectral contrasts for F2 and F3 was significant only for the 1.0, 1.5, and 2.0 cycles/octave filter conditions ($p<0.05$). Given that normal-hearing listeners needed a minimum of 3 dB F1 and F2 spectral contrasts to identify synthetic vowels with 90% accuracy (Leek *et al.*, 1987), the reduction in F1 and F2 spectral contrasts is consistent with the decrease in vowel identification with increasing spectral modulation filtering.

3. Spectral tilt

Figure 2 also shows a reduction in spectral tilt with increasing the spectral modulation filter cutoff frequency. Spectral tilt was computed for unfiltered and filtered vowels by calculating the dB/octave difference between F1 and F3 peak amplitudes following Kiefte and Kluender (2005). As shown in Fig. 8, spectral tilt was negative for the unfiltered vowels and approached or exceeded zero as the filter cutoff frequency increased to 2 cycles/octave. A one-way (filter) ANOVA indicated a significant effect of high-pass spectral modulation filter cutoff on spectral tilt ($F_{4,44}$=49.773, $p<0.05$). A Tukey *post hoc* test revealed that the spectral tilt for unfiltered vowels was significantly more negative than for vowels filtered at 0.5, 1.0, 1.5, and 2.0 cycles/octave (all p values <0.05).

4. Correlation between acoustic features and vowel identification

Correlational analyses were conducted to assess the potential relationships among vowel identification scores and various local acoustic features modified as a result of filtering in the spectral modulation domain including local F1–F3

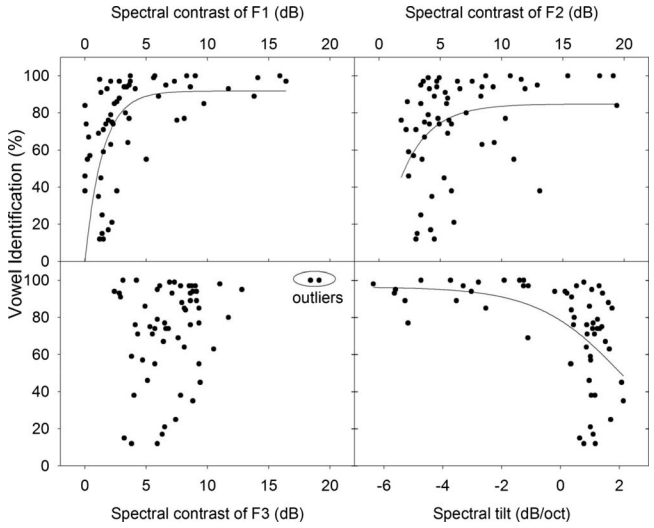


FIG. 9. Scatter plots and nonlinear exponential regression functions showing the relationship between vowel identification averaged over five listeners and four acoustic features computed from the vowel stimuli including F1 contrast (upper left), F2 contrast (upper right), F3 contrast (lower left), and spectral tilt (lower right). The exponential fit for F3 is not shown due to the lack of a significant regression.

spectral contrasts as well as spectral tilt. There were modest but significant correlations between vowel identification and each acoustic feature (see Fig. 9, all $|r|<0.5$; all $p<0.05$) except the spectral contrast of F3 when the two obvious outlying points (see oval in the lower left panel of Fig. 9) were removed ($p=0.135$). Given that vowel identification scores ranged from 0% to 100%, nonlinear exponential regression, rather than linear regression, was then undertaken to reveal the detailed relationship between vowel identification and acoustic features. Nonlinear regressions, based on the 12 vowel categories and 5 filter conditions (i.e., unfiltered and four high-pass filtered conditions), were computed with vowel identification scores, averaged over five listeners, as the dependent variable and four acoustic features, including the spectral contrast associated with F1–F3 as well as spectral tilt, as independent variables. The two outlying data points associated with F3 contrast again were removed for the regression analysis. As shown in Fig. 9, regressions based on an exponential fit were significant for the spectral contrast associated with F1 and F2 and the spectral tilt variables (all $p<0.05$), but not for the spectral contrast associated with F3. These results indicate that the degradation in vowel identification for this female talker was associated with reductions in F1 and F2 contrasts and increases in spectral tilt that resulted from filtering in the spectral modulation domain.

IV. DISCUSSION

The present study investigated the acoustic and perceptual impact of attenuating the low spectral modulation components of vowels by progressively increasing the high-pass cutoff frequency of a simple Butterworth high-pass filter applied in the spectral modulation frequency domain. The results showed that the presence of spectral-modulation frequencies below 2 cycles/octave was important in preserving the basic spectral features of vowels that are required for identification (i.e., spectral contrast and spectral tilt). Attenuating spectral modulation frequencies up to 2 cycles/octave resulted in a dramatic and progressive reduction in vowel identification performance, especially for back vowels. In addition to reducing F1 and F2 spectral contrasts and increasing spectral tilt, the effect of these spectral changes produced a more crowded vowel space.

American English is one of the world's languages with a relatively crowded vowel space, typically characterized by considerable spectral overlap in the two-dimensional F1 $\times$F2 vowel formant space (Peterson and Barney, 1952; Hillenbrand *et al.*, 1995). Studies of clear speech indicate that the broader F1$\times$F2 vowel space of clear speech leads to significant improvement in the identification of isolated vowels by normal-hearing and hearing-impaired listeners (Ferguson and Kewley-Port, 2002). As shown in Fig. 6, the manipulation of spectral modulation filter cutoff frequency in this study produced a more crowded vowel space, consistent with greater confusion among vowels, especially back vowels. Indeed, the four back vowels, /ɔ, o, u, U/, were more closely spaced in the F1$\times$F2 vowel space for the 2 cycles/octave than the unfiltered condition, consistent with the perceptual confusions among these back vowels reported in Tables I and II.

Another important acoustic feature that was substantially affected by spectral modulation filtering was the spectral contrast associated with the vowel formants. As shown in Fig. 7, increases in the high-pass cutoff frequency resulted in a decrease in the spectral contrast that varied across F1–F3. Averaging across the three formants, the spectral contrast ranged from 9.8 (no filtering) to 3.3 dB (2 cycles/octave filter). Several previous studies directly manipulated the spectral contrast corresponding to formant frequencies and reported vowel identification performance ranging from 75% to 82% correct for contrast values of between 1 and 3 dB (Summerfield *et al.*, 1987; Leek *et al.*, 1987; Alcantara and Moore, 1995). By comparison, average vowel identification (across 12 vowels) for the 0.5 cycle/octave filter condition reported here was 78.5% correct and corresponded to an average formant spectral contrast of 7.8 dB. Clearly, the spectral contrast required to produce ~78% correct was much higher in the present study relative to the previous studies cited. There are several methodological and acoustic differences between the present and previous studies that may contribute to differences in the contrast associated with approximately 78% correct performance, perhaps the most important of which are the number of vowel categories (12 versus 4, 5, or 6), the stimulus type (resynthesized natural speech versus uniform harmonic complexes with formantlike peaks), and the fact that changes in addition to spectral contrast were present in the current study. Specifically, the spectral envelopes of the 12 vowels in the present study were highly irregular, whereas the spectral envelopes in the previous studies were flat with the exception of equal-amplitude increments in the two harmonics closest to the F1–F3 frequencies. This distinction in some respects is similar to the detection of spectral features in low versus high uncertainty

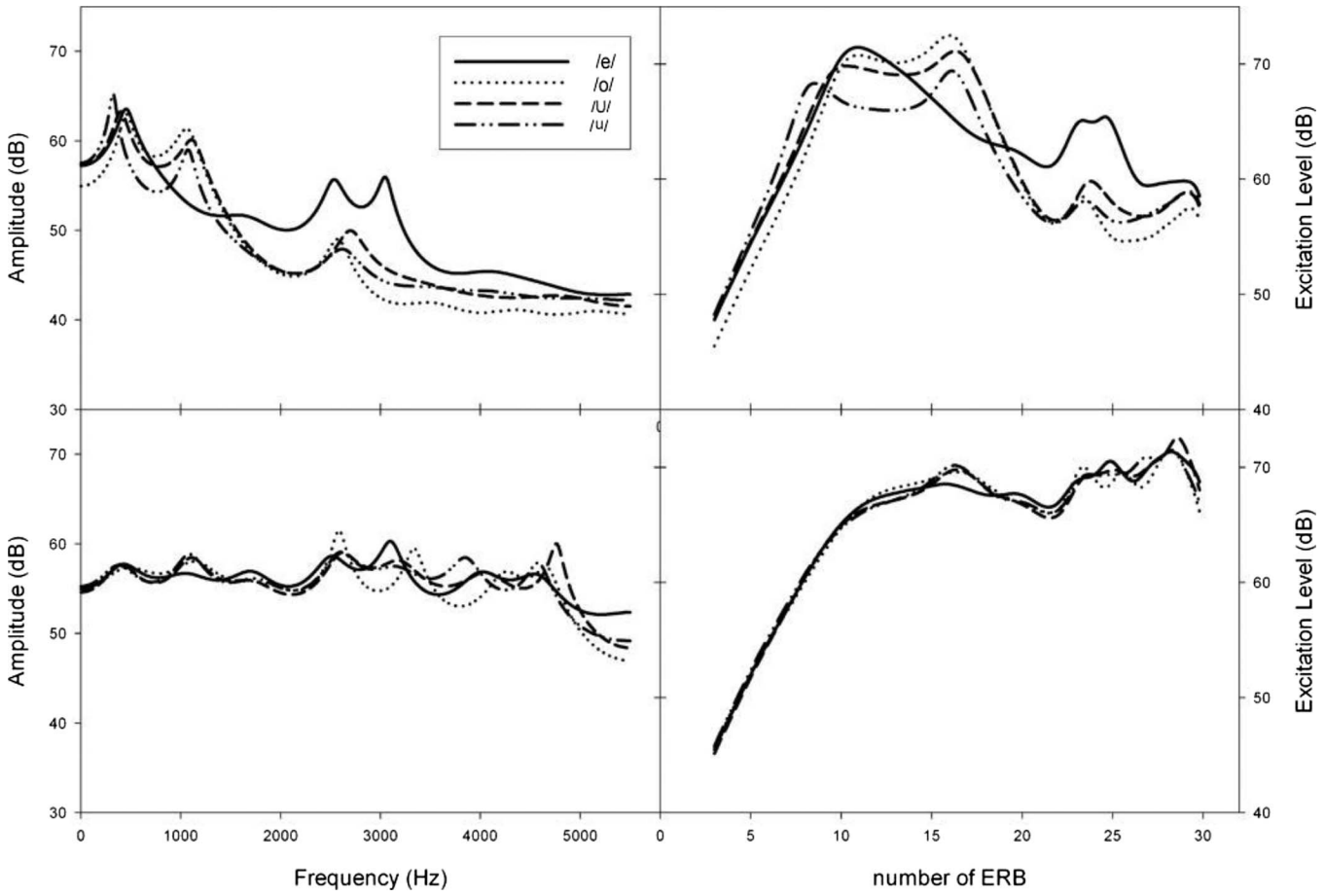


FIG. 10. LPC spectra (left) and excitation patterns (right) for the unfiltered vowels (top), /e, ɑ, o, u/, and the same four vowels filtered at 2 cycles/octave (bottom).

conditions (e.g., Neff and Green, 1987), where several spectral features contributed to the uncertainty in the present task.

Figure 10 shows the LPC spectra (left) and excitation patterns (right) of four vowels, /e, ɑ, o, u/, without filtering and high pass filtered at 2 cycles/octave. Excitation patterns for the four vowels were calculated following Moore and Glasberg's procedure (1987). Formant peaks F1 and F2 are well represented in the excitation patterns for unfiltered vowels (upper right panel) but are poorly represented in the excitation patterns for filtered vowels (lower right panel), with F1 and F2 merging into one spectral prominence. Thus, compared to the internal representation of the unfiltered vowels, high-pass filtering at 2.0 cycles/octave resulted in very similar excitation patterns for the four vowels. This is consistent with the fact that the four vowels were highly confused with each other in the 2.0 cycles/octave condition, even though the vowel /e/ was relatively distant from the other three vowels in the F1 × F2 vowel space (see Fig. 5). These analyses also highlight the fact that analyses of the internal representation of various stimuli may provide much more information than analyses of the acoustic features themselves. For the three back vowels /ɑ, o, u/ filtered at 2.0 cycles/octave, the acoustic analyses above demonstrated that the small spectral contrasts and closely spaced F1 and F2 peaks are not preserved by the peripheral transformations of the auditory system emulated in the excitation pattern computations. Rather, those peaks are merged into single spectral prominences that differ little among the vowels. Similar comparisons of the LPC spectra (left) and excitation patterns (right) for the four front vowels, /æ, ɛ, i, ɪ/, without filtering and high pass filtered at 2 cycles/octave are shown in Fig. 11. The excitation patterns of the filtered vowels /æ/ and /ɛ/ are quite similar to each other, consistent with the perceptual measures indicating that these two vowels were highly confused with each other (see Table I, lower section). Unlike the back vowels that showed merged F1 and F2 peaks in the excitation patterns (see Fig. 10), both F1 and F2 peaks for the four front vowels were fairly well preserved in the excitation patterns for the 2 cycles/octave condition even though the spectral contrasts were substantially reduced (lower right panel of Fig. 11). This is consistent with the perceptual measures indicating relatively high intelligibility (>75% on average) for the front vowels even in the 2 cycles/octave condition.

The decreased vowel identification associated with the reduction in spectral contrast reported here is consistent with previous studies of spectral smearing (ter Keurs *et al.*, 1992, 1993; Drullman *et al.*, 1996), although the techniques that led to reduced spectral contrast were quite different in these studies. In the experiments reported by ter Keurs *et al.* (1992, 1993), the spectral envelope was convolved with a Gaussian-shaped filter of fixed relative bandwidth, whereas the study of Drullman *et al.* (1996) produced spectral smearing by either low-pass or high-pass filtering the temporal envelope.

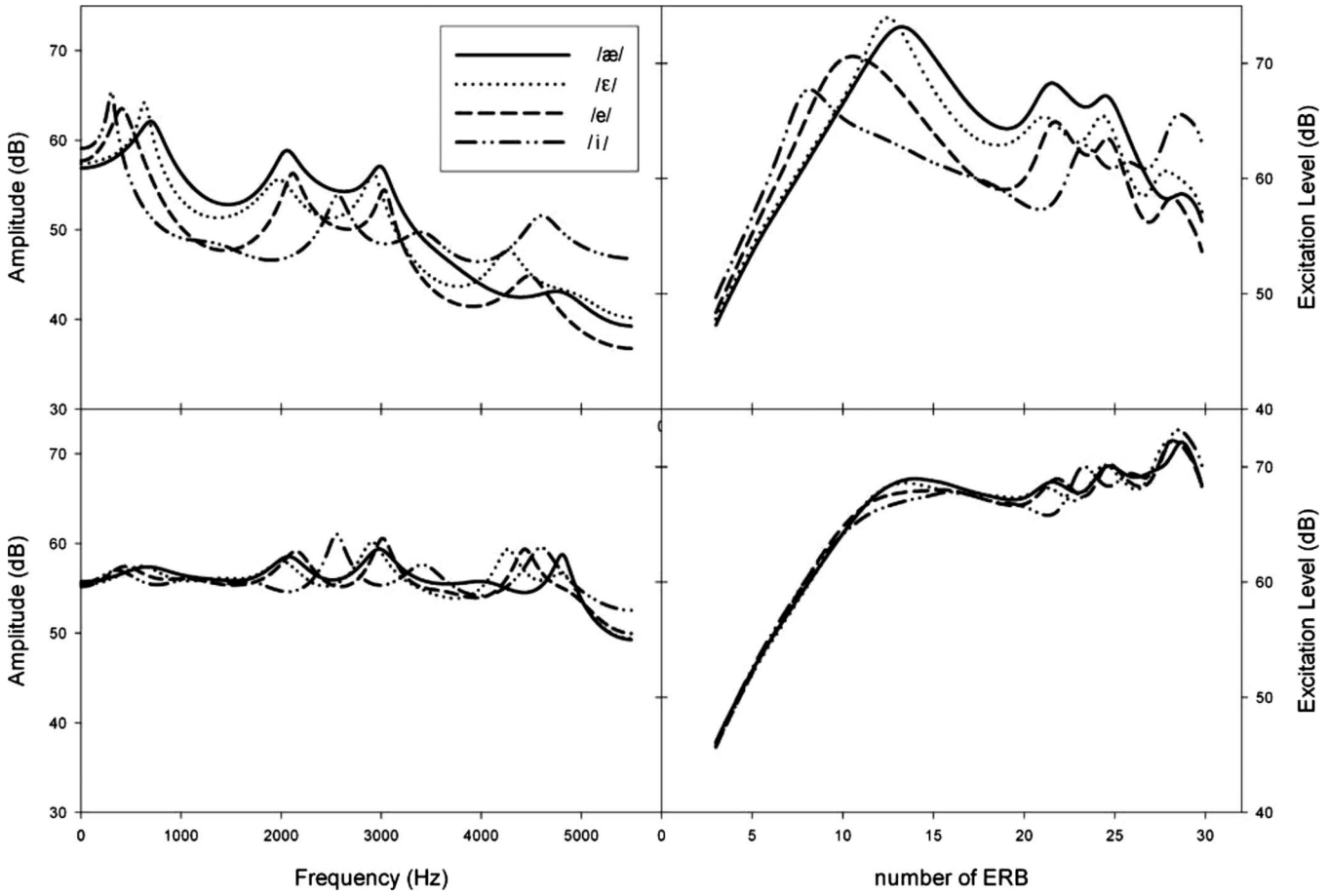


FIG. 11. LPC spectra (left) and excitation patterns (right) for the unfiltered front vowels (top), /i, e, ɛ, æ/, and the same four vowels filtered at 2 cycles/octave (bottom).

The fact that the different acoustic manipulations of all these studies resulted in a reduction in spectral contrast and an associated decrease in vowel identification indicates that spectral contrast, especially for F1 and F2, is a critical factor in vowel perception.

Spectral tilt is another acoustic feature that was significantly altered by spectral modulation filtering. As described in the Introduction, spectral tilt significantly affects vowel quality (Ito *et al.*, 2001; Kiefte and Kluender, 2005), especially when the time-varying acoustic features of the vowels are not preserved (e.g., Kiefte and Kluender, 2005). Because the formant and fundamental frequencies of the vowels were time invariant in this study, the increased spectral tilt for the filtered vowels relative to the unfiltered vowels may have contributed to the decrease in vowel identification with increasing filter cutoff frequency. This is consistent with the relationship between vowel identification and spectra tilt, as shown in Fig. 9. For each of the three vowel groups: front, central, and back vowels, vowel identification was significantly correlated with spectral tilt ($r=-0.422$ for front, -0.797 for central, and -0.703 for back, $p<0.05$), indicating the importance of spectral tilt in vowel perception. In addition, as explained above, spectral tilt differed by about 100% between the original front and back vowels, but differed by only 10% between the front and back vowels for the 2 cycles/octave filter condition. Such changes in spectral tilt were associated with the finding that back vowels were confused with front vowels much more frequently (35.4%) for the two cutoff filter condition than for the unfiltered condition (0%; see Tables I and II), consistent with the interpretation that spectral tilt is a contributing factor in the discrimination of front versus back vowels. These results were consistent with Kiefte and Kluender (2005) who reported that the identification of synthetic vowels with greater (i.e., more negative) spectral tilt was biased toward back vowels relative to front vowels. For example, using synthetic vowels with F1 and F2 frequencies characteristic of a typical /u/ vowel, they systematically varied the spectral tilt from that of a typical /u/ vowel to that of a typical /I/ vowel (i.e., from a steep negative slope to a shallower slope). Identification score of the /u/ vowel was approximately 10% greater for the typical /u/ spectral tilt than for the typical /I/ spectral tilt (see their Fig. 4), illustrating the use of spectral tilt as a cue for vowel identity.

In summary, the global spectral shape of English vowels was modified by high-pass filtering in the spectral modulation domain. Acoustic analyses comparing unfiltered and filtered vowels revealed consistent reductions in spectral contrast, increased spectral tilt, and a slightly smaller F1×F2 vowel space. These changes in the global spectral shape were associated with perceptual changes marked by dramatic reductions in vowel identification performance, indicating that global spectral features are critical for vowel identification (Zahorian and Jagharghi, 1993; Blandon and Lindblom,

1981). Furthermore, these results demonstrate that systematic changes in the spectral modulation domain lead to systematic changes in both vowel identification and a set of local acoustic parameters that are individually related to vowel identification. Conversely, these results indicate that spectral modulation frequencies below 2 cycles/octave are important for preserving the spectral features of vowels that contribute to accuracy in vowel identification. The extent to which listeners with poor vowel identification performance (e.g., hearing-impaired listeners) exhibit reductions in both spectral shape perception and corresponding changes in similar local spectral features is unknown but currently under study. The demonstration of a systematic relationship between parametric manipulations in the spectral modulation domain below 2.0 cycles/octave and vowel identification performance leads to the interesting possibility that gain within the same range of spectral modulation frequencies might improve vowel perception, particularly in the presence of competing sounds or for listeners who have reduced frequency selectivity and/or spectral shape perception.

ACKNOWLEDGMENTS

The authors are grateful to two anonymous reviewers and Dr. Mitchell Sommers for their constructive and detailed comments on the earlier version of the manuscript. Research was supported by NIH NIDCD R01 DC04403 and NIA P01 AG09524.

[1]Spectral modulation frequency quantifies the relative distance between spectral features such as intensity peaks (e.g., formants). Changes in the modulation depth (spectral contrast) at several spectral modulation frequencies, as a consequence of high-pass filtering in the spectral modulation domain, for example, can result in changes in the distance among spectral peaks. In terms of vowel stimuli, this can result in formant shifts.

Alcantara, J., and Moore, B. (**1995**). "The identification of vowel-like harmonic complexes: Effects of component phase, level, and fundamental frequency," J. Acoust. Soc. Am. **97**, 3813–3824.
ANSI (**2004**). "Specification for Audiometers," ANSI Report No. S3.6-2004, ANSI, New York.
Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.
Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.
Bladon, R. A. W. (**1982**). "Arguments against formants in the auditory representation of speech," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. CrantrÖm (Elsevier Biomedical, Amsterdam), pp. 95–102.
Blandon, R. A. W., and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.
Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech perception," J. Acoust. Soc. Am. **95**, 1053–1064.
Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of reducing slow temporal modulations on speech perception," J. Acoust. Soc. Am. **95**, 2670–2680.
Drullman, R., Festen, J. M., and Houtgast, T. (**1996**). "Effect of temporal modulation reduction on spectral contrasts in speech," J. Acoust. Soc. Am. **99**, 2358–2364.
Eddins, D. A., and Bero, E. M. (**2007**). "Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region," J. Acoust. Soc. Am. **121**, 363–372.
Ferguson, S. H., and Kewley-Port, D. (**2002**). "Vowel intelligibility in clear and conversational for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.
Fu, Q., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.
Hillenbrand, J., Getty, L. J., Clark, M. J., and Weeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hillenbrand, J. M., and Nearey, T. (**1999**). "Identification of resynthesized /hVd/ utterances: effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.
Ito, M., Tsuchida, J., and Yano, M. (**2001**). "On the effectiveness of whole spectral shape for vowel perception," J. Acoust. Soc. Am. **110**, 1141–1149.
Kawahara, H., Masuda-Kastuse, I., and Cheveigne, 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.
Kiefte, M., and Kluender, K. (**2005**). "The relative importance of spectra tilt in monophthongs and diphthongs," J. Acoust. Soc. Am. **117**, 1395–1404.
Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 148–154.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liu, C., and Fu, Q. (**2007**). "Estimation of vowel recognition with cochlear implant simulations," IEEE Trans. Biomed. Eng. **54**, 74–81.
Loizou, P. (**2000**). CILEA: A Matlab software tool for speech analysis available: http://www.utdallas.edu/~loizou/speech/colea.htm (accessed 8/15/07).
Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2134.
Moore, B. C. J., and Glasberg, B. R. (**1987**). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," Hear. Res. **28**, 209–225.
Moore, B. C. J., Vickers, D. A., Glasberg, B. R., and Baer, T. (**1997**). "Comparison of real and simulated hearing impairment subjects with unilateral and bilateral cochlear hearing loss," Br. J. Audiol. **31**, 227–245.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.
Peterson, G. E., and Barney, H. L. (**1952**). "Control method used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Qian, J., and Eddins, D. A. (**2008**). "The role of spectral cues in virtual sound localization revealed by spectral modulation filtering and enhancement," J. Acoust. Soc. Am. **123**, 302–314.
Summerfield, A. Q., Sidwell, A., and Nelson, T. (**1987**). "Auditory enhancement of changes in spectral amplitude," J. Acoust. Soc. Am. **81**, 700–708.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech reception. I," J. Acoust. Soc. Am. **91**, 2872–2880.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Effect of spectral envelope smearing on speech reception. II," J. Acoust. Soc. Am. **93**, 1547–1552.
Turner, C. W., Chi, S. L., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.
van Veen, T. M., and Houtgast, T. (**1985**). "Spectral sharpness and vowel dissimilarity," J. Acoust. Soc. Am. **77**, 628–634.
Zahorian, S., and Jagharghi, A. (**1993**). "Spectral-shape feature versus formant as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.

Linguistic experience and audio-visual perception of non-native fricatives

Yue Wang[a)]
Department of Linguistics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

Dawn M. Behne
Department of Psychology, Norwegian University of Science and Technology, N7491 Trondheim, Norway

Haisheng Jiang
Department of Linguistics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

(Received 7 December 2007; revised 13 June 2008; accepted 13 June 2008)

This study examined the effects of linguistic experience on audio-visual (AV) perception of non-native (L2) speech. Canadian English natives and Mandarin Chinese natives differing in degree of English exposure [long and short length of residence (LOR) in Canada] were presented with English fricatives of three visually distinct places of articulation: interdentals nonexistent in Mandarin and labiodentals and alveolars common in both languages. Stimuli were presented in quiet and in a café-noise background in four ways: audio only (A), visual only (V), congruent AV (AVc), and incongruent AV (AVi). Identification results showed that overall performance was better in the AVc than in the A or V condition and better in quiet than in café noise. While the Mandarin long LOR group approximated the native English patterns, the short LOR group showed poorer interdental identification, more reliance on visual information, and greater AV-fusion with the AVi materials, indicating the failure of L2 visual speech category formation with the short LOR non-natives and the positive effects of linguistic experience with the long LOR non-natives. These results point to an integrated network in AV speech processing as a function of linguistic background and provide evidence to extend auditory-based L2 speech learning theories to the visual domain.
 [DOI: 10.1121/1.2956483]

PACS number(s): 43.71.Hw [AJ] Pages: 1716–1726

I. INTRODUCTION

Language experience often involves face-to-face interaction with simultaneous perception of a speaker's voice and facial movements. For native (L1) perceivers speech perception is enhanced with visual information (e.g., Erber, 1969; Sumby and Pollack, 1954), especially when auditory distinctiveness decreases, such as in a noisy environment (Bernstein *et al.*, 2004; Erber, 1969; Sumby and Pollack, 1954; Summerfield, 1979). The relative contribution of audio and visual information has also been revealed by what is known as the "McGurk effect," where an audio [ba] dubbed onto a visual [ga] may produce a [da] percept (McGurk and MacDonald, 1976). These results suggest an ability to integrate auditory and visual speech information (Massaro, 1987, 1998; Sekiyama *et al.*, 2003). A question that arises is whether this ability reflects an innate capacity to process multimodal information or is developed by learning and experience such that the information processing is instantiated by language-specific patterns.

A hardwired ability to process audio and visual information has been shown for prelinguistic infants: 2–4 month olds prefer speech information matching, compared to non-matching, in auditory and visual information (Kuhl and Meltzoff, 1982; Patterson and Werker, 1999, 2003) and can integrate auditory and visual speech information with incongruent input modalities (Burnham and Dodd, 2004; Desjardins and Werker, 2004; Rosenblum *et al.*, 1997). However, while 4–6-month-old infants can visually distinguish their L1 speech from an unfamiliar language, 8 month olds cannot (Weikum *et al.*, 2007), suggesting a shift from language-universal to language-specific multimodal speech processing as infants are exposed to a specific language. Cross-linguistic studies with adults have also shown language specific, as well as language-universal patterns. For example, while Chinese, English, Japanese, and Spanish perceivers demonstrate similar audio-visual processing patterns when perceiving a [ba]-[da] continuum (Chen and Massaro, 2004; Massaro *et al.*, 1993), Chinese perceivers show worse visual perception of [ba] and [da] than Dutch perceivers (de Gelder and Vroomen, 1992). Similarly, although the McGurk effect has consistently been observed across languages (Sams *et al.*, 1998; Sekiyama and Tohkura *et al.*, 1993; Werker, Frost, and McGurk, 1992), native speakers of some languages such as Chinese and Japanese (Sekiyama, 1997; Sekiyama and Tohkura, 1993) show a weaker magnitude of this effect than do English or Spanish perceivers (Massaro *et al.*, 1993). It has been speculated that Chinese speakers' relatively diminished use of visual speech information may be due to Chinese being a tonal language and thereby relying on the visually less distinct prosodic information (Sekiyama, 1997; Sekiyama and Tohkura, 1993), although significant visual in-

a)Author to whom correspondence should be addressed. Electronic mail: yuew@sfu.ca

formation for tones in Chinese has also been observed (Burnham *et al.*, 2001). More recent research has nevertheless demonstrated that Mandarin Chinese perceivers can use the visual speech information in their L1 to the same degree as English perceivers do (Chen and Hazan, 2007a). These discrepancies suggest the need for further cross-linguistic research to address the issue of language specificity of audio-visual speech processing.

The perception of non-native (L2) speech provides a unique case. On the one hand, visual speech information when available may enhance the perception and production of an L2 (Massaro *et al.*, 1993; Massaro, 1998). For example, with the addition of visual information, English perceivers demonstrate more accurate perception of French (Reisberg *et al.*, 1987), Korean (Davis and Kim, 2004), Irish, and Spanish sounds (Erdener and Burnham, 2005). Similarly, Spanish-dominant Spanish-Catalan bilinguals can distinguish the Catalan vowel contrast [e-ɛ] in an audio-visual condition but not with audio-only or visual-only input (Navarra and Soto-Faraco, 2007). For Korean and Japanese learners of English, perception of non-native [f] and [l] improves with matched auditory and visual input compared to audio-visual mismatched stimuli (Hardison, 1999), and the identification of the most difficult auditory stimuli receives the greatest benefit from visual cues (Hardison, 2005a).

Nevertheless, non-natives are less efficient at using visual information than natives (de Gelder and Vroomen, 1992; Hazan *et al.*, 2005, 2006; Ortega-Llebaria *et al.*, 2001; Sekiyama, 1997). For example, unlike native Spanish and Catalan perceivers, Italian and English natives cannot discriminate Catalan and Spanish when only visual information is available (Soto-Faraco *et al.*, 2007). Non-natives, such as L2 perceivers of English and Japanese, appear to be more affected by the McGurk illusion for stimuli presented in an L2 than in an L1 (Sekiyama and Tohkura, 1993). Exceptions to this are Chinese perceivers showing an equally weak McGurk effect for their L1 and L2 (Japanese or English), suggesting language-specific differences based on a perceiver's L1 (Sekiyama, 1997; Sekiyama, Tohkura, and Umeda, 1996). Notably, in these studies with native Japanese and Chinese perceivers, all the target stimuli involve audio-visual phonemes common to both the L1 and L2, which may not correspond to perceptual patterns of new L2 audio-visual stimuli. Indeed, visual information would more likely be beneficial for phonemes occurring both in L1 and L2, whereas the addition of visual information would not as easily reduce confusion for visual cues specific to new L2 phones (Ortega-Llebaria *et al.*, 2001). For example, Spanish learners show more sensitivity than Japanese learners to the visual cues to English [v], because the visual category for the labiodental [v] is allophonically existent in Spanish as the voiceless counterpart [f], whereas it is totally missing in Japanese (Hazan *et al.*, 2006). These findings indicate that L2 learners have difficulty in correctly using the visual cues that do not contrast phonetically in their L1.

The patterns of audio-visual perception are reminiscent of the L2 speech learning theories developed on the basis of auditory perception [e.g., speech learning model (SLM), Flege, 1995; the perceptual assimilation model (PAM), Best, 1995), stating that L2 learners' inability to accurately perceive L2 sounds is due to incorrect assimilation of L2 sounds with L1 phonetic categories and therefore failure in establishing new L2 phonetic categories (Best *et al.*, 1988; Best, 1995; Flege, 1987, 1995). In this respect, the acquisition of L2 visual categories (termed "visemes," Brooke and Summerfield, 1983; Walden *et al.*, 1977) may be comparable to that of auditory L2 sound learning (Erdener and Burnham, 2005; Hardison 1999, 2003; Hazan *et al.*, 2005, 2006). L2 perceivers may have lost their sensitivity to visual cues that are not used in their L1 and need to learn to associate these visual cues with a viseme in the L2 (Hazan *et al.*, 2005, 2006) to establish new L2 categories.

Visual speech learning may occur with linguistic experience. Sensitivity to visual information in L2 sounds can be enhanced through audio-visual training, with the degree of improvement affected by one's L1 (Hardison, 2003, 2005b; Hazan *et al.*, 2005). Research has also shown a positive correlation between overall L2 proficiency and proficiency in visual L2 speech perception. Werker *et al.* (1992) examined the visual perception of incongruent audio-visual English fricatives by native French perceivers with varying English proficiency and found that for the interdental fricative absent in French, French speakers' degree of fluency in English was positively related to their correct use of the English visual information.

In sum, these findings indicate that the difficulty in L2 visual perception of speech may lie in visual speech categories that are not used in the L1, suggesting that subsequent research should focus on new L2 visual categories. Furthermore, the presence of cross-linguistic differences also suggests the need for further research to include additional L2 learners (such as Chinese learners who have been found to use less visual information in their L1) to examine possible language-specific differences. In addition, if linguistic experience plays a role in L2 visual perception, further studies are necessary to identify the kinds of linguistic experience that may be facilitative. Moreover, nonlinguistic factors may also affect L2 audio-visual processing. For example, given that L1 visual speech perception is particularly effective in non-optimal conditions, such as in noise (e.g., Sumby and Pollack, 1954; Summerfield, 1979), how noise affects the perception of new L2 visual speech information will further illuminate the role of language background in audio-visual processing.

A. The current study

On the basis of these previous findings, the current research examines auditory and visual perception of the English interdental fricatives in quiet and noisy backgrounds by native Mandarin Chinese perceivers varying in length of residence (LOR) in Canada.

Firstly and primarily, this study focuses on the visual perception of interdental fricatives and hypothesizes that English interdentals nonexistent in Mandarin perceivers' L1 are more difficult to perceive than the place-adjacent labiodentals and alveolars familiar to them in Mandarin. Previous research on Mandarin natives' perception of speech visemes

TABLE I. Participants' language background information. [LOR: mean length of residence in Canada (years); AOA: mean age of arrival in Canada; AOL: age of initial English learning; English study: mean number of years of formal English study; Mandarin input: mean percent daily input in Mandarin (L1); English input: mean percent daily input in English (L2). Means are based on averages across participants, and the range across participants is included in parentheses.]

	Age (yrs)	Gender (F/M)	LOR (yrs)	AOA (yrs)	AOL (yrs)	English study (yrs)	Mandarin input (%)	English input (%)
English ($n=15$)	22 (19–28)	9F/6M	…	…	…	…	…	…
Mandarin long LOR ($n=15$)	21 (18–25)	10F/5M	10 (8–13)	11 (9–13)	11 (9–13)	12 (10–16)	38 (25–63)	62 (38–75)
Mandarin short LOR ($n=20$)	24 (19–25)	15F/5M	2 (1–4)	21 (18–23)	11 (8–12)	12 (7–17)	60 (25–100)	40 (0–75)

common to their L1 and L2 (e.g., Sekiyama *et al.*, 1996, Sekiyama, 1997) may not reflect the perception of new L2 visemes. Since the difficulty of L2 visual speech perception lies in the new L2 visual speech categories (Hazan *et al.*, 2005, 2006), interdental fricatives are used in the current study. Furthermore, since Mandarin perceivers demonstrate less use of visual speech information in their L1 compared to perceivers of other languages (e.g., Burnham *et al.*, 2001; de Gelder and Vroomen, 1992; Sekiyama *et al.*, 1996; Sekiyama, 1997), questions arise as to how this affects their visual perception of L2 sounds.

The second focus is the effect of L2 experience indexed by the LOR in an L2 environment. LOR, which has been used as an index in L2 auditory speech learning research (e.g., Flege, 1995; 1998; Flege *et al.*, 1995; Flege *et al.*, 1999; McAllister *et al.*, 2002; Riney and Flege, 1998), may be particularly relevant for visual speech perception since L2 learners typically do not have extensive exposure to L2 visual cues until they arrive in an L2 country. Based on findings that L2 visual perception improves with short-term training (Hardison, 2003; Hazan *et al.*, 2005) and overall L2 proficiency (Werker *et al.*, 1992), the current study hypothesizes that the long LOR Mandarin perceivers outperform the short LORs in perceiving the English visemes nonexistent in Mandarin. Furthermore, differences observed between the two Mandarin groups will inform patterns of visual speech learning in progress.

Additionally, this study addresses the influence of background noise. For native perceivers, the contribution of visual input is especially effective in a noisy environment (Bernstein *et al.*, 2004; Sumby and Pollack, 1954; Summerfield, 1979). L2 research on the auditory domain shows that non-native listeners are more affected by noise than native listeners, with a reduction in intelligibility when speech is degraded in noise (Hazan *et al.*, 2005; Mayo and Florentine, 1997; Sekiyama *et al.*, 2003). On this basis, L2 perceivers may particularly need to rely on visual speech information. However, if they cannot effectively pick up visual speech information in the L2, their performance would be similar in quiet and noisy backgrounds. In other words, if audio-visual processing is not language-specific, noise would be expected to affect non-native and native perceptions in the same way. Alternatively, language-specific audio-visual processing would result in different processing patterns in L1 and L2 as a function of listening backgrounds.

II. METHODS

A. Participants

Three groups of young adults (mean age=22) participated in the study: one group of native Canadian English participants ($n=15$), and two groups of native Mandarin Chinese differing in their LOR in Canada (long LOR: $n=15$; short LOR: $n=20$). The participants' background information is summarized in Table I. The "long LOR" Mandarin group consisted of native Mandarin English users who had resided in the country for an average of 10 years. They moved to Canada, on average, at age 11, when they were initially exposed to English. All reported having English as their dominant language since their arrival and having attended Canadian secondary schools where English was the language of education. The Mandarin "short LOR" group had an average of two years' residency in Canada. Before arriving in Canada, they had studied English as an L2 in a classroom setting (5 hours/week) since they were on average 12 years old and had not resided in any other English-speaking environments prior to their arrival in Canada. Based on their reported language background, they continued to use Mandarin as their more dominant language since their arrival. These selection criteria ensure a comparable age range between the two Mandarin groups at the time of testing (21–24 years old), comparable age of initial English learning (11–12 years old), and comparable length of English learning (12 years). All three groups reported having normal or corrected vision and no known history of speech and hearing impairments. All were undergraduate or graduate students at Simon Fraser University (SFU), Canada at the time of the study. They were compensated for their participation.

B. Stimuli

Stimuli were based on 18 English consonant-vowel (CV) syllables having a fricative onset followed by a vowel: [fi, fa, fu, vi, va, vu, θi, θa, θu, ði, ða, ðu, si, sa, su, zi, za,

zu]. The fricatives differed in place of articulation (labiodental, interdental, alveolar), representing a sequence of audio-visual categories involving articulators from more front to further back in the vocal tract. While all three places of articulation are phonemically distinctive in English, the labiodental and alveolar fricatives occur in Mandarin (as voiceless), whereas the interdentals do not occur in Mandarin. The voiced and voiceless counterparts of each syllable were included to allow a broad test of place of articulation identification. The vowels ([i, a, u]) occur in Mandarin as well as in English, representing a range of vocal tract configurations varying in tongue height, advancement, and lip rounding (Hazan *et al.*, 2006; 2005; Jongman *et al.* 2003). Thus each target fricative place of articulation was represented by six different exemplars, to ensure responses to phonetic categories rather than acoustic idiosyncrasies.

On this basis, stimuli were developed which had (1) congruent audio and visual components (AVc), (2) only an audio component (A), (3) only a video component (V), and (4) incongruent (mismatched) audio and visual components (AVi) (see McGurk and MacDonald, 1976). In the AVi stimuli, the fricative place of articulation was an audio-visual combination of labiodental and alveolar ($[A_{labiodental}-V_{alveolar}]$, e.g., audio [fa] dubbed onto video [sa], $A_{fa}-V_{sa}$; or $[A_{alveolar}-V_{labiodental}]$, e.g., audio [sa] dubbed onto video [fa], $A_{sa}-V_{fa}$). If AV-fusion occurs, the percept is expected to have a place of articulation intermediate to labiodental and alveolar, which is the interdental place non-native to Mandarin perceivers (e.g., [θa]). The fricative voicing and vowel were always the same for the A and V components of a given AVi syllable (e.g., $A_{fa}-V_{sa}$, $A_{fu}-V_{su}$, and $A_{va}-V_{za}$). A total of 12 AVi stimuli were used (two AV input place $[A_{labiodental}-V_{alveolar}, A_{alveolar}-V_{labiodental}]\times$ two voicing conditions [voiceless, voiced] × three vowels [i, a, u]).

1. Production

Audio and video recordings were made in the Language and Brain Lab at SFU with an adult male speaker of Canadian English producing six repetitions of the 18 syllables (six fricatives × three vowels) at a normal speaking rate. Video recordings of the speaker's face were made using a digital camcorder (Sony DCR-HC30/40) placed about 3 m from the speaker against a white background. Separate audio recordings were simultaneously made with a Shure KSM 109 condenser microphone to a PC via an audio interface (M-audio MobilePre USB) at a 44.1 kHz sampling rate. These high quality audio recordings were used to replace the audio track from the camcorder recording.

2. Editing

The video recordings with their camcorder audio recordings were transferred to a PC. From among the six repetitions of each syllable, a best example of each syllable was selected such that the durational difference among the 18 selected syllables was under 10%, the approximate just noticeable difference for duration (Lehiste, 1970).

For these selected syllables, the video recordings were aligned with corresponding high quality audio recordings using SOUNDFORGE 8.0. This was done by synchronizing the waveforms of the syllable recorded from the camcorder and the Shure microphone. The audio signal from the camcorder was then deleted. The final set of audio stimuli was normalized so that all the resulting stimuli had the same unweighted rms value. Visual stimuli were edited such that each clip had a 1.2 s neutral face before and 1.2 s after the stimulus, with the length of the stimulus measured from the frame where mouth movement was first detected (mouth opening) to the frame where the mouth movement ended and the mouth was fully closed. The frame length was 0.06 s, and the resolution was 640 × 480 pixels.

The resulting AV stimuli were used as AVc stimuli and as the basis for creating the A-only, V-only, and AVi stimuli. The A and V sets of stimuli were created from the AVc files by removing the video tracks or muting the audio tracks, respectively. For the AVi stimuli, the A and V components were the same as those used in the AVc, A-only, and V-only conditions. Each AVi stimulus was created by aligning audio and video components from different syllables (e.g., $A_{fa}-V_{sa}$): starting from the AVc stimulus, which contained the desired V component (e.g., AVc [sa]), the audio signal of the desired A component (e.g., [fa]) was synchronized to the onset of the fricative audio signal of the AVc and the original AVc audio (e.g., audio [sa]) was deleted, so that it was replaced by the new audio signal (e.g., [fa]).

3. Goodness evaluation

Stimulus goodness evaluation tests were carried out on the final sets of AVc and AVi stimuli to confirm the intelligibility of the audio signals and the naturalness of the audio-visual signals, each with two phonetically trained native speakers of English. Intelligibility of the audio signals was tested with an identification task, showing 95% correct responses for the audio signals in the AVc stimuli and 100% correct responses for the audio signals in the AVi stimuli. The naturalness of the AV stimuli was determined by asking the evaluators whether the audio and video signals were naturally synchronized, regardless of what they heard or saw. This was carried out with a 5-point goodness rating task (5 being the most natural), showing that the AVc stimuli were rated 4.4 and the AVi stimuli rated 4.6.

4. Stimulus background

Finally, cafeteria noise recorded at SFU was added to a copy of all the stimuli, resulting in a quiet and café-noise (S/N = 0 dB) version for all the syllables. The signal-to-noise (S/N) ratio was established empirically by testing two English listeners at each of the following S/N values: 10, 0, and −10 dB. A S/N ratio of 0 dB was found to generate an error rate of about 30%. At 10 dB, identification was similar to the quiet background condition, whereas at −10 dB, signals, especially the fricatives, were completely masked. Therefore, cafeteria noise at 0 dB S/N ratio was embedded in each of the CV syllables, with the duration of the noise corresponding to the syllables' duration, to create the set of stimuli in the noisy condition (A-noise, V-noise, AVc-noise, AVi-

TABLE II. Mean percent correct responses for the main effects in the identification of the target fricative stimuli by English and Mandarin long LOR and short LOR participants, as a function of (i) place of articulation, (ii) modality, and (iii) background. Standard deviations are in parentheses.

	Place of articulation			Modality			Background	
	Labiodental	Interdental	Alveolar	A	V	AVc	Quiet	Noise
English	86 (3.2)	78 (3.8)	73 (2.5)	82 (2.0)	64 (2.8)	91 (2.4)	84 (2.0)	73 (2.1)
Mandarin long LOR	84 (3.2)	64 (3.8)	65 (2.5)	73 (2.0)	56 (2.8)	84 (2.4)	76 (2.0)	65 (2.1)
Mandarin short LOR	77 (2.8)	54 (3.3)	59 (2.2)	63 (1.7)	54 (2.4)	73 (2.1)	69 (1.7)	57 (1.8)

noise). In total, 132 stimuli were used, including 66 in quiet (18 A, 18 V, 18 AVc, 12 AVi) and 66 in a café-noise background.

C. Procedure

The perception experiment was generated in E-PRIME 1.0 (Psychology Software Tools), allowing stimulus presentation (integrating video clips imported from Microsoft Powerpoint files) and response log. An identification task was carried out where each participant was tested on the full set of stimuli in the quiet and noisy backgrounds. Stimuli were blocked by modality (A, V, and AV) and background (quiet and noise), with AVi and AVc stimuli in the same block. Each of these six blocks included two repetitions of each stimulus. Stimuli were randomized within a block. The order of background and modality presentation was counterbalanced across participants.

For each trial a stimulus was presented auditorily over loudspeakers, visually on a computer monitor, or both. The participants were tested individually, seated about 1 m from a 20 in. liquid crystal display (LCD) flat panel computer monitor and two loudspeakers (Altec Lansing) positioned directly at each side of the monitor, such that the sound and image source had approximately the same distance to the perceivers' ears and eyes. Loudspeakers rather than headsets were used to prevent perceivers from being biased toward the audio input. The audio signal was presented at a comfortable level of approximately 70 dB, found to be sufficient to achieve an optimal score (Nábelek and Robinson, 1982; Takata and Nábelek, 1990).

For each trial, a fixation point was displayed in the center of the monitor for 1 s, followed by the target stimulus. Response alternatives were presented on the monitor, where, for a given trial, the alternatives were the six fricatives [f, v, θ, ð, s, z], as well as the option to type in an alternative response. The symbols "th" and "dh" were used to represent [θ] and [ð], respectively (Jongman *et al.*, 2003). The participants' task was to identify the fricative and respond by pressing a key on the computer keyboard for the corresponding symbol displayed on the monitor. Participants had up to 4 s to respond.

The test was preceded by a short introduction session containing task instructions, stimulus familiarization (e.g., matching the symbols and the sounds they represent), and five practice trials for each modality (A, V, and AV). The test ended with a debriefing session during which participants filled out a post-test questionnaire. The whole test lasted about 1 h, including a short break after three test blocks.

III. RESULTS

Data were analyzed separately for the audio, visual, and audio-visual congruent conditions (A, V, and AVc), and for the audio-visual incongruent condition (AVi).

A. Audio, visual, and congruent audio-visual conditions

The percent correct identification of the fricatives was analyzed using a four-way mixed analysis of variance (ANOVA), with group (English, Mandarin long LOR, and Mandarin short LOR) as a between-subject factor, and place of articulation (POA) (labiodental, interdental, and alveolar), modality (A, V, and AVc), and background (quiet and noise) as repeated measures. The dependent variable was the perceivers' correct identification of POA regardless of consonant voicing and vowel context (Hazan *et al.*, 2005, 2006; Jongman *et al.*, 2003; Werker *et al.*, 1992) since POA was the focus of interest.

Table II presents mean percentage of correct responses for the English, Mandarin long, and short LOR groups as a function of POA, modality, and background. A significant main effect of group was found [$F(2,47)=17.6$, $p<.001$], with the *post hoc* analysis (Tukey HSD),[1] showing that native English perceivers had a higher overall percentage of correct responses than the Mandarin long LOR group, which in turn had a higher percentage of correct responses than the Mandarin short LORs. Significant main effects were also observed for POA [$F(2,47)=30.2$, $p<.001$], modality [$F(2,47)=183.1$, $p<.001$], and background [$F(1,47)=381.1$, $p<.001$]. Across participant groups, identification was better with labiodentals than with interdentals or alveolars, better in the AV than in the A or V condition, and better in quiet than in noise.

Moreover, native group differences were observed with significant interactions of group × modality [$F(4,47)=2.9$, $p=.027$], group × POA × modality [$F(8,47)=2.5$, $p=.013$], and group × POA × modality × background [$F(8,47)=6.1$, $p<.001$]. Further analyses were thus conducted to compare group differences for POA, modality, and background. Figure 1 displays the results.

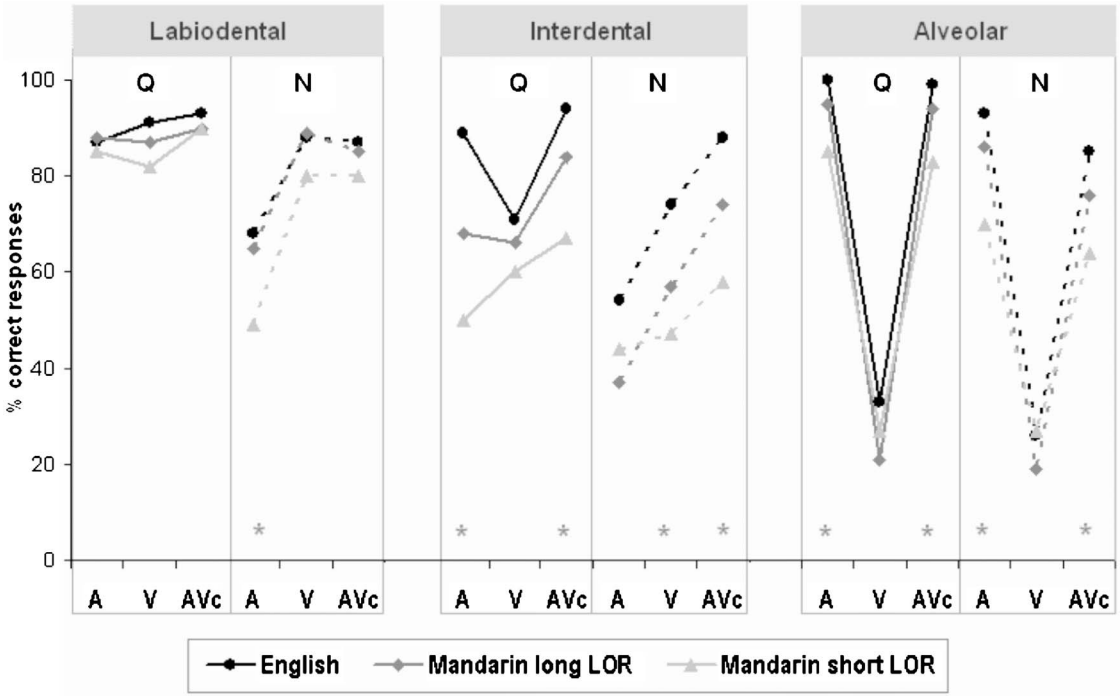


FIG. 1. Mean percent correct responses for each place of articulation (labiodental, interdental, and alveolar) and modality (A, V, and AVc) in quiet (Q, left panels) and noise (N, right panels) backgrounds by English (black lines) and Mandarin Chinese long LOR group (dark gray lines) and short LOR group (light gray lines). " *" represents significant ($p<.05$) statistical results comparing among the three participant groups for each place of articulation and input modality.

First, sets of two-way (modality and group) repeated measure ANOVAs were conducted for each POA in quiet to compare how the AV modalities were perceived as a function of group. Only interdental identification showed a modality and group interaction [$F(4,44)=4.0$, $p=.005$]. *Post hoc* analyses show that whereas the English natives' identification in the V condition was poorer than the A and AVc conditions, the Mandarin long LOR group's A and V identifications were poorer than AVc identification, and the short LOR group did poorer in the A (than V than AVc) modality. To compare how different fricative POAs were perceived, sets of two-way ANOVAs were carried out in each modality in quiet. A POA and group interaction was observed in the A [$F(4,44)=6.8$, $p<.001$] and AV [$F(4,44)=3.4$, $p<.001$] conditions. In the A condition, whereas the English natives' identification of both the interdentals and labiodentals was moderately lower than for the alveolars, both the Mandarin long and short LOR groups' identification for the interdentals was significantly lower than for the labiodentals and alveolars. For the AV stimuli, while the native English and Mandarin long LOR groups did not differ in their performance for the three POAs, the Mandarin short LOR group had a much lower percentage of correct responses for the interdentals than labiodentals and alveolars.

To compare group differences directly, sets of one-way ANOVAs were carried out with group as a between-subject factor in each modality and POA in quiet. Differences were observed for the interdentals and alveolars in the A and AVc conditions. For the interdentals, the A condition revealed a decreasing identification accuracy from the native English to Mandarin long LOR to short LOR groups [$F(2,47)=23.6$, $p<.001$], and in the AVc condition, both the English and the Mandarin long LOR group's accuracy was higher than that of the short LOR group's [$F(2,47)=8.3$, $p<.001$]. For the alveolars, both the native English and Mandarin long LOR group's percentages of correct responses were higher than the Mandarin short LOR group's [A: $F(2,47)=10.1$, $p<.001$; AVc: $F(2,47)=14.0$, $p=.001$]. The V condition, however, did not reveal a significant group difference for interdentals (all being moderately low) or alveolars (all being low). For the labiodentals, no difference was observed across modalities, all of which were relatively high.

Given that most of the between group differences occurred with interdentals, confusion patterns of interdental identification were analyzed (Table III). In both A and AV conditions, the Mandarin short LOR group misperceived 18–19% of the interdentals as alveolars, while the English natives never confused these two POAs in quiet. Similarly, in the V condition, the short LOR group was also more biased toward the alveolars (30%) than toward the labiodentals (9%), while the English natives misperceived the interdentals as labiodentals or alveolars to a similar degree. The Mandarin long LOR group generally followed the native patterns. It is also noted that although the participants were given the option to type in an alternative to the six fricatives, this "other" option was not much used (1%–3% across conditions and participant groups).

For stimuli in noise, cross-modality comparisons show that significant group differences only occurred with interdentals: while an increase in accuracy from A-noise to V-noise to AVc-noise was observed for both the English natives [$F(2,42)=12.1$, $p<.001$] and the Mandarin long LOR group [$F(2,42)=8.6$, $p=.001$], the short LOR group only

TABLE III. Confusion matrix for the A, V, and AVc conditions. Mean percent responses (%) in the identification of the interdental fricatives given the response options: labial, interdental, and alveolar by English and Mandarin long LOR and short LOR participants.

		Interdental stimuli								
		English			Mandarin long LOR			Mandarin short LOR		
	Perceived as	A	V	AVc	A	V	AVc	A	V	AVc
Quiet	Labiodental	11	11	6	31	16	14	31	9	13
	Interdental	89	71	94	68	65	84	50	60	67
	Alveolar	0	16	0	1	19	2	19	30	18
Noise	Labiodental	24	14	7	13	13	10	11	15	10
	Interdental	54	74	88	37	57	74	44	47	58
	Alveolar	22	11	4	49	29	16	42	37	30

showed a marginal difference across modalities $[F(2,57)=3.1,\ p=.051]$, all being low. Direct comparisons of the noise and quiet conditions for each POA and modality show that while the perception was generally poorer in noise than in quiet ($p<.05$), the Mandarin short LOR group's auditory perception of interdentals did not differ between the noise and quiet conditions $[F(1,19)=1.5,\ p=.240]$, both being relatively low.

B. AV incongruent condition

For each AVi stimulus, corresponding responses were tabulated based on whether the consonant in the response matched the consonant in the audio component of the stimulus (A-match), the video component (V-match), or the fused A and V components (AV-fusion, i.e., interdental). Sets of three-way mixed ANOVAs were carried out for each of these response types, with group as a between-subject factor and AV-place input ($[A_{labiodental}+V_{alveolar}]$, $[A_{alveolar}+V_{labiodental}]$) and background as repeated measures. Figure 2 displays the mean percent responses for each response type as a function of AV place input in quiet and noisy backgrounds.

Results for AV-fusion showed a significant effect of AV-place $[F(1,47)=31.3,\ p<.001]$ and a group×AV-place interaction $[F(1,47)=6.3,\ p=.004]$. A moderate degree of

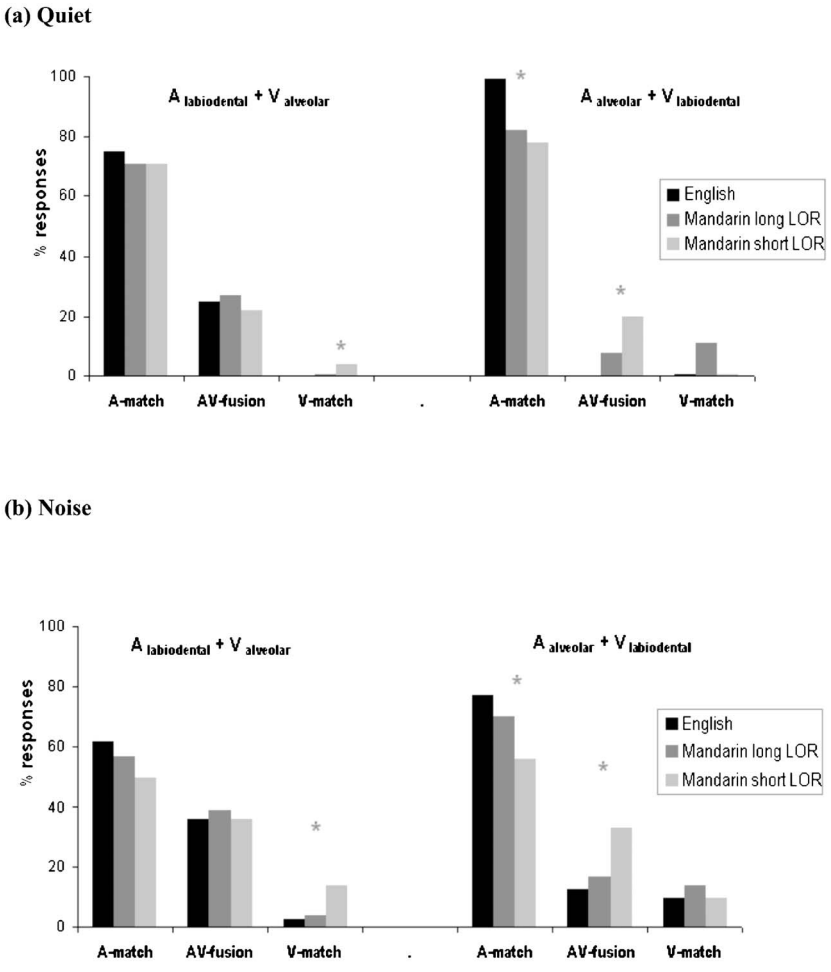


FIG. 2. Mean percent responses for incongruent AV stimuli in (a) quiet and (b) noise backgrounds by English (black bars), Mandarin Chinese long LOR group (dark gray bars), and short LOR group (light gray bars). "A match" and "V match" indicate correct responses matching A and V components, respectively. "AV-fusion" refers to interdental responses matching neither the A nor the V component, corresponding to the McGurk effect. "*" represents significant ($p<.05$) statistical results comparing among the three participant groups for each place of articulation and input modality.

fused interdental responses (25%) was observed across groups for the $A_{labiodental}+V_{alveolar}$ condition where one would expect the McGurk effect. However, in the $A_{alveolar}+V_{labiodental}$ input condition, the Mandarin short LOR group had a higher mean percent than the long LOR group whose responses were in turn greater than the English group [$F(2,47)=12.2$, $p<.001$], indicating that the Mandarin perceivers more easily fused the incongruent stimuli despite the intermediate fricatives being non-native. Furthermore, a reliable effect was observed across groups for background [$F(1,47)=55.3$, $p<.001$], with greater fusion in noise than in quiet.

For percent responses matching the A component with $A_{alveolar}+V_{labiodental}$ stimuli, the English group had a higher mean score than the Mandarin long LOR group, which in turn was higher than the short LOR group [quiet: $F(2,47)=5.1$, $p=.001$] [noise: $F(2,47)=4.6$, $p=.015$], indicating that the native group was more accurate in using the auditory component. For the V-match responses, on the other hand, a group difference was observed with $A_{labiodental}+V_{alveolar}$ stimuli, where in both quiet and noise backgrounds, the short LOR group scored higher than long LOR and English perceivers [quiet: $F(2,47)=5.03$, $p=.010$], [noise: $F(2,47)=3.9$, $p=.027$].

C. Summary of results

Results across participant groups show identification in general to be better with AV stimuli than with A- or V-only stimuli and better in quiet than in noise. Responses to the incongruent AV stimuli were dominated by the A input, but the likelihood to use V information increased in noise. Significant group differences were nonetheless evident, particularly with the interdental fricatives nonexistent in Mandarin, showing decreasing identification accuracy from the native English to Mandarin long LOR to short LOR groups. While the English natives' interdental identification in the A condition was poorer than that in the V condition only in noise, the Mandarin short LOR group showed poorer use of A than V information in quiet as well as in noise. Although both the Mandarin long and short LOR groups' performances improved in the AVc condition, only the long LOR group achieved the native level. In the incongruent condition, the Mandarin perceivers, particularly the short LOR group, were more likely to give an AV-fused interdental response than the English perceivers. These results indicate that the native English group relied more on the A information, with the V information coming in to facilitate perception in noise. The Mandarin short LOR group, on the other hand, appears to have made more use of V than A information in the perception of the non-native interdentals. The results also reveal the effect of LOR, with the Mandarin long LOR group's performance being intermediate to that of the English and Mandarin short LOR groups.

IV. DISCUSSION

A. Native and universal patterns

Consistent with previous findings of similar consonant and fricative perception (Chen and Hazan, 2007a, 2007b; Hardison, 1999; Jongman *et al.*, 2003; Massaro, 1998; Sekiyama and Tohkura, 1993; Werker *et al.*, 1992), results show that the English natives can accurately perceive the fricatives in quiet in the A-only and AVc condition. That the perception (especially of the interdentals and alveolars) is poorer in the V-only than in the other conditions also agrees with previous research (Chen and Hazan 2007a, 2007b; Cienkowski and Carney, 2002; McGurk and McDonald, 1976). Similarly, as has been reported (Erber, 1969; Hardison, 1999; Summerfield, 1979), native English perceivers are able to effectively adopt the V cues when the signal is masked in noise.

It has been argued that some aspects of AV speech perception may be neutral across languages (Chen and Massaro, 2004; Massaro, 1998). The present findings show that the native and non-native groups all actively use the audio component, especially in the incongruent AV condition and in the perception of alveolars, which as sibilants are auditorily (acoustically) more robust than the nonsibilant labiodentals and interdentals. On the other hand, all groups use visual cues when available, but reveal poor performance for the visually presented sounds that are not visually distinct (e.g., alveolars), suggesting that a visual benefit is in part determined by the visual salience of the place of articulation (e.g., Dodd, 1977; Hardison, 1999; Hazan *et al.*, 2006). Additionally, according to the fuzzy logical model of perception (FLMP) (Massaro, 1998), although the input information may differ for specific languages, the operational processes of AV perception are universal across languages, similar to those for the processing of low-level nonlinguistic auditory and visual information. Thus, visual speech information may similarly affect perceivers of different languages, with the degree of its influence being largest to the extent that the audio information is ambiguous (Chen and Massaro, 2004). In the present results, supporting evidence is shown by the similarity between the AV perception of non-native sounds (by Mandarin perceivers) and that of the native sounds in noise (by English perceivers), where perceivers tend to pay more attention to the visual speech input in nonoptimal listening conditions, be it involving new non-native auditory information, or in a poor listening environment.

B. Non-native patterns

The non-native perception of labiodentals and alveolars has a general pattern similar to the English natives', supporting that visemes common to the L1 and L2 do not present problems in L2 perception (e.g., Hardison, 1999; Hazan *et al.*, 2005, 2006; Werker *et al.*, 1992).

Of particular interest is the Mandarin perceivers' identification of the non-native interdentals. It has been reported that native Mandarin perceivers use less visual speech information in their L1 compared to perceivers of other languages, thus the prediction that Mandarin perceivers may not be sensitive to the L2 visual speech information (Burnham *et al.*, 2001; de Gelder and Vroomen, 1992; Hazan *et al.*, 2005; Sekiyama *et al.*, 1996; Sekiyama, 1997). The current results, however, show that Mandarin perceivers' (especially the short LOR group's) correct interdental identification in

the V-only condition is greater than that in the A-only condition, and they show a greater use of V information and greater magnitude of AV-fusion in the incongruent condition than native English perceivers. These results suggest that the Mandarin perceivers made greater use of L2 visual speech information, despite not weighing the visual input heavily in their L1. Indeed, previous research has shown that the perception of L2 stimuli improves with additional visual information (Hardison, 1999) and that visual cues enhance L2 speech comprehension (Navarra and Soto-Faraco, 2007; Reisberg *et al.*, 1987; Soto-Faraco *et al.*, 2007). Consistently, Japanese perceivers who do not weigh visual speech information heavily in their L1 also demonstrate more AV-fusion in perceiving English than in perceiving Japanese sounds (Sekiyama *et al.*, 1996). Given that perceivers rely more on visual speech information when auditory intelligibility is poor (Erber, 1969; Sumby and Pollack, 1954; Summerfield, 1979), L2 perceivers conceivably resort to the visual information as an additional channel of input in perceiving the difficult non-native sounds (Hattori, 1987; Hardison, 2003).

The Mandarin perceivers' interdental identification is nevertheless poor across input modalities, leading to the critical question of whether non-native perceivers simply make greater use of visual information or if they can adopt the L2 specific visual cues. For the incongruent condition, the Mandarin (short LOR) perceivers have more occurrences of AV-fusion (with the fused interdentals being non-native) than the English perceivers, indicating that non-natives are more vulnerable to the AV illusion. Even though they make greater use of the visual input, they cannot effectively use these visual cues in a linguistically meaningful manner. In most of the earlier studies with similar results (e.g., Burnham and Dodd, 1998; Chen and Hazan, 2007a, 207b; Sekiyama and Tohkura, 1993; Sekiyama *et al.*, 2003), the fused sound often has had a place of articulation existent in perceivers' L1; thus the present results extend these findings showing consistent patterns for an AV-fused (interdental) sound nonexistent in the perceivers' L1. Moreover, the natives but not the (short LOR) non-natives make more fused responses for $A_{labiodental}+V_{alveolar}$ stimuli than $A_{alveolar}+V_{labiodental}$ stimuli. This direction effect has also been reported previously, with fusion more easily occurring when the visual input is not visually salient (McGurk and McDonald, 1976; Sekiyama and Tohkura, 1991). However, that this effect in the current study decreases from the natives to the long and short LOR groups suggests that the non-natives are less sensitive to the difference of the L2 visual input. Overall, the results suggest that although L2 perceivers tend to use as much information as possible to compensate for the difficulty in the perception of non-native sounds, awareness of the visual speech domain does not necessarily lead to an accurate perception of L2 visual cues.

C. Linguistic experience

Extending previous findings showing a facilitative role of experience on L2 AV perception (e.g., Hardison, 2003; Sekiyama, 1997; Werker *et al.*, 1992), the results reveal the effect of linguistic exposure indexed by the length of residence in an L2 country. Compared to the Mandarin short LOR group, the long LOR group can more correctly identify the non-native interdentals and integrate the AV information, as well as being less susceptible to the AV-fusion. They approximate the native English patterns with a greater degree of accuracy in perceiving auditory than visual information, but with visual information becoming particularly facilitative in noise when the listening condition is poor. Together, these results reveal a pattern of learning in progress. As less experienced perceivers' (the short LOR group) auditory perception of L2 sounds is poor, they focus on visual information as an additional channel of input, yet cannot use the visual information correctly or efficiently integrate the auditory and visual information. In contrast, experienced non-native perceivers (the long LOR group) make a nativelike use of AV information. These results not only reveal that AV perception of L2 sounds changes as a function of linguistic experience, but also suggest a dynamic learning pattern whereby auditory learning may precede and is accompanied by visual learning, resulting in an effective integration of AV speech information.

D. Noise background

A comparison of the native and non-native perceivers' performance in quiet and noise reveals that a significant group difference exists only in the interdental identification. Visual information facilitates the English natives' perception in noise, as they can efficiently weigh the audio and visual input channels based on need, relying more on visual input in a nonoptimal listening condition. The Mandarin (short LOR) natives, however, show poor audio perception in both quiet and noise, and the addition of visual information does not facilitate perception to the same degree as for the natives. These patterns indicate that noise differentially affects native and non-native AV perception for the L2 sounds, suggesting a language-specific AV processing.

E. General discussion

Evidence from the present study points to language-specific processing integrated with universal aspects of AV processing. The results of interdental perception suggest language-specific patterns whereby native and non-native speakers weigh the auditory and visual input differently. While native English perceivers are primarily dependent on the auditory component, non-native perceivers (the short LOR group in particular) tend to make greater use of the visual information, although their performance remained poor in the manner in which they used the L2 visual cues and in effectively integrating the L2 AV speech information. While previous research indicates that L2 perceivers do not use visual information to the same degree as natives (e.g., Hazan *et al.*, 2006; Ortega-Llebaria *et al.*, 2001; Sekiyama, 1997; Werker *et al.*, 1992), this study further shows that focusing on the visual speech domain does not necessarily lead to an accurate perception. Furthermore, that the Mandarin long LOR group outperformed the short LOR group indicates that L2 visual speech cues may be learned and improved with L2 experience. Despite these language-specific

patterns, the similarity between the native and non-native visual perception with degraded auditory information (as in noise, and when non-native, respectively) also suggests universal aspects in the underlying integration of AV speech information (consistent with the FLMP, Chen and Massaro, 2004).

Particularly promising from the present results along with previous L2 AV research is the possibility of bridging the learning patterns across speech input modalities. Indeed, the acquisition of non-native visual cues is assumed to be analogous to that of auditory L2 speech learning (Hazan *et al.*, 2005, 2006; Ortega-Llebaria *et al.*, 2001; Sekiyama, 1997). The difficulty in the perception of L2 visual information may be attributed to the interference from an L1. In particular, L2 visual cues may be classified as "identical," "similar," or "new," depending on whether they have gestural counterparts in the L1, as has been proposed in L2 auditory speech learning theories (e.g., SLM, Flege, 1995; PAM, Best, 1995). For L2 visual information which is non-assimilable to an L1 category, learners need to learn to associate L2 specific visual cues to corresponding L2 phones in order to establish new L2 categories. In the present visual perception results, the short LOR Mandarin perceivers predominantly confuse the interdentals with the place-adjacent alveolars familiar in their L1 (see Table III), indicating that L2 category formation may be blocked not only if an L2 phone is auditorily perceived as similar to the closest L1 category (Flege, 2007), but also if L2 and L1 phones are visually close. Recent theories also suggest that in addition to the visual similarity between L1 and L2, L2 speech learning models should include further factors to account for audio-visual L2 speech processing, such as the relative weighing of auditory and visual cues and the distinctiveness of visual cues (Hazan *et al.*, 2006). The non-native short LOR group may have increased visual weighing in the perception of the new L2 interdentals but still have difficulty assimilating them to the appropriate L2 categories. On the other hand, as is the case with the acquisition of auditory speech, the long LOR group's nativelike performance indicates the formation of new L2 visual categories with exposure, suggesting a similar role of linguistic experience on L2 speech perception across auditory and visual domains.

V. CONCLUDING REMARKS

The language-specific as well as universal aspects of AV processing revealed in this paper and in previous research suggest an integrated network in cognitive processing and learning that involves and goes beyond individual modalities and domains. Results also suggest that the L2 speech learning theories developed on the basis of auditory speech perception may be extended to include visual speech perception.

These findings point to directions for future research. First, the similarities and dissimilarities of L1 and L2 visual/gestural categories deserve to be further explored and eventually quantified. Subsequent research should include more extensive types of L2 speech sounds that characterize various visemic relationships with corresponding L1 sounds (e.g., identical, similar, and new). Second, the present study reveals the effect of linguistic experience indexed by length of residence in the L2 environment. As previous research suggests that other factors (such as age of L2 acquisition, amount of L2 input, and L2 proficiency) may vary and covary to influence auditory speech perception (e.g., Flege, 1995, 1998; Flege *et al.*, 1999; McAllister *et al.*, 2002), these factors should also be taken into account in L2 visual speech learning. Finally, the present findings may have pedagogical implications: that is, training and/or teaching of L2 speech should not only emphasize the awareness of the visual domain (to increase visual weighting), but should also direct learners to focus on L2 specific visual speech cues (to establish correct L2 categories).

ACKNOWLEDGMENTS

We thank Nicole Carter, Angela Cooper, Chad Danyluck, Angela Feehan, Vivian Hsing, Nina Leung, Elaine Pang, Lindsay Shaw, and Kristy Stefanucci at Simon Fraser University (SFU) for their assistance in stimulus development, data collection, analysis, and preparation of the manuscript. We would also like to thank Dr. Allard Jongman and the three anonymous reviewers for their valuable suggestions. Portions of this research were presented at Interspeech 2006 in Pittsburgh and at the Fourth Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan in Honolulu in 2006. This project was supported by a standard research grant from the Social Sciences and Humanities Research Council of Canada (SSHRC 410-2006-1034) and a SFU Institutional SSHRC grant.

[1]Here and elsewhere Tukey HSD was carried out for *post hoc* analyses where there are three levels of a factor. The accepted significance level was $p < .05$.

Bernstein, L. E., Auer, E., and Takayanagi, S. (**2004**). "Auditory speech detection in noise enhanced by lipreading," Speech Commun. **44**, 5–18.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (Baltimore, York), pp. 171–204.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 45–60.

Brooke, N. M., and Summerfield, Q. (**1983**). "Analysis, synthesis, and perception of visible articulatory movements," J. Phonetics **11**, 63–76.

Burnham, D., and Dodd, B. (**1998**). "Familiarity and novelty in infant cross-language studies: Factors, problems, and a possible solution," in *Advances in Infancy Research*, edited by C. Rovee-Collier, pp. 170–187.

Burnham, D., and Dodd, B. (**2004**). "Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk Effect," Dev. Psychobiol. **45**, 204–220.

Burnham, D., Lau, S., Tam, H., and Schoknecht, C. (**2001**). "Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers," in Proceedings of the International Conference on Auditory-Visual Speech Processing, edited by D. Massaro, J. Light, and K. Geraci, pp. 155–160.

Chen, Y., and Hazan, V. (**2007a**). "Developmental Factor in Auditory-Visual Speech Perception—The McGurk Effect in Mandarin-Chinese and English Speakers," in Proceedings of the International Conference on Auditory-Visual Speech Processing, edited by J. Vroomen, M. Swerts, and E. Krahmer.

Chen, Y., and Hazan, V. (**2007b**). "Language effects on the degree of visual influence in audiovisual speech perception," in Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrueken, Germany, pp. 2177–2180.

Chen, T., and Massaro, D. W. (**2004**). "Mandarin speech perception by ear and eye follows a universal principle," Percept. Psychophys. **66**, 820–836.
Cienkowski, K. M., and Carney, A. E. (**2002**). "Auditory-visual speech perception and aging," Ear Hear. **23**, 439–449.
Davis, C., and Kim, J. (**2004**). "Audio-visual interactions with intact clearly audible speech," Q. J. Exp. Psychol. **57A**, 1103–1121.
de Gelder, B., and Vroomen, J. (**1992**). "Auditory and visual speech perception in alphabetic and non-alphabetic Chinese/Dutch bilinguals," in *Cognitive Processing in Bilinguals*, edited by R. J. Harris (Elsevier, Amsterdam), pp. 413–426.
Desjardins, R., and Werker, J. F. (**2004**). "Is the integration of heard and seen speech mandatoryfor young infants?," Dev. Psychobiol. **45**, 187–203.
Dodd, B. (**1977**). "The role of vision in the perception of speech," Perception **6**, 31–40.
Erber, N. P., (**1969**). "Interaction of audition and vision in the recognition of oral speech stimuli," J. Speech Hear. Res. **12**, 423–425.
Erdener, V., and Burnham, D. (**2005**). "The role of audiovisual speech and orthographic information in nonnative speech production," Lang. Learn. **55**, 191–228.
Flege, J. E. (**1987**). "A critical period for learning to pronounce foreign languages?," Appl. Linguist. **8**, 162–177.
Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience*, edited by W. Strange (Baltimore, York), pp. 233–273.
Flege, J. E. (**1998**). "The role of subject and phonetic variables in L2 speech acquisition," in *Papers from the 34th Annual Meeting of the Chicago Linguistics Society, Volume II, The Panels*, edited by M. Gruber, D. Higgins, K. Olsen, and T. Wysocki (Chicago Linguistic Society, Chicago), pp. 213–232.
Flege, J. E. (**2007**). "Language contact in bilingualism: Phonetic system interactions," in *Laboratory Phonology 9*, edited by J. Cole and J. I. Hualde (Mouton de Gruyter, Berlin).
Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995**). "Factors affecting strength of perceived foreign accent in a second language," J. Acoust. Soc. Am. **97**, 3125–3134.
Flege, J. E., Yeni-Komshian, G., and Liu, S. (**1999**). "Age constraints on second language learning," J. Mem. Lang. **41**, 78–104.
Hardison, D. M. (**1999**). "Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment," Lang. Learn. **49**, 213–283.
Hardison, D. M. (**2003**). "Acquisition of second-language speech: Effects of visual cues, context, and talker variability," Appl. Psycholinguist. **24**, 495–522.
Hardison, D. M. (**2005a**). "Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment," Appl. Psycholinguist. **26**, 579–596.
Hardison, D. M. (**2005b**). "Variability in bimodal spoken language processing by native and nonnative speakers of English: A closer look at effects of speech style," Speech Commun. **46**, 73–93.
Hattori, T. (**1987**). "A study of nonverbal intercultural communication between Japanese and Americans—Focusing on the use of the eyes," Japan Assoc. Lang. Teachers **8**, 109–118.
Hazan, V., Sennema, A., Faulkner, A., and Ortega-Llebaria, M. (**2006**). "The use of visual cues in the perception of non-native consonant contrasts," J. Acoust. Soc. Am. **119**, 1740–1751.
Hazan, A., Sennema, A., Iba, M., and Faulkner, A. (**2005**). "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English," Speech Commun. **47**, 360–378.
Jongman, A., Wang, Y., and Kim, B. (**2003**). "Contributions of semantic and facial information to perception of nonsibilant fricatives," J. Speech Lang. Hear. Res. **46**, 1367–1377.
Kuhl, P. K., and Meltzoff, A. N. (**1982**). "The bimodal perception of speech in infancy," Science **218**, 1138–1141.
Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).
Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Erlbaum, Hillsdale, NJ).
Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT, Cambridge, MA).
Massaro, D. W., Tsuzaki, M., Cohen, M. M., Gesi, A., and Heredia, R. (**1993**). "Bimodal speech perception: An examination across language," J. Phonetics **21**, 445–478.
Mayo, L. H., and Florentine, M. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.
McAllister, R., Flege, J., and Piske, T. (**2002**). "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English, and Estonian," J. Phonetics **30**, 229–258.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Nábelek, 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.
Navarra, J., and Soto-Faraco, S. (**2007**). "Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds," Psychol. Res. **71**, 4–12.
Ortega-Llebaria, M., Faulkner, A., and Hazan, V. (**2001**) "Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English," in Proceedings of the International Conference on Auditory-Visual Speech Processing, edited by D. Massaro, J. Light, and K. Geraci, pp. 149–154.
Patterson, M. L., and Werker, J. F. (**1999**). "Matching phonetic information in lips and voice is robust in 4.5-month-old infants," Infant Behav. Dev. **22**, 237–247.
Patterson, M. L., and Werker, J. F. (**2003**). "Two-month-old infants match phonetic information in lips and voice," Dev. Sci. **6**, 191–196.
Reisberg, D., McLean, J., and Goldfield, A. (**1987**). "Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli," in *Hearing by Eye: The Psychology of Lip-reading*, edited by B. Dodd and R. Campbell (Erlbaum, London), pp. 97–113.
Riney, T. J., and Flege, J. E. (**1998**). "Changes over time in global foreign accent and liquid identifiability and accuracy," Stud. Second Lang. Acquis. **20**, 213–244.
Rosenblum, L. D., Schmuckler, M. A., and Johnson, J. A. (**1997**). "The McGurk effect in infants," Percept. Psychophys. **59**, 347–357.
Sams, M., Manninen, P., Surakka, V., Helin, P., and Katto, R. (**1998**). "McGurk effect in Finnish syllables, isolated words, and words in sentences: effects of word meaning and sentence context," Speech Commun. **26**, 75–87.
Sekiyama, K. (**1997**). "Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects," Percept. Psychophys. **59**, 73–80.
Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (**2003**). "Auditory-visual speech perception examined by fMRI and PET," Neurosci. Res. (N Y) **47**, 277–287.
Sekiyama, K., and Tohkura, Y. (**1991**). "McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility," J. Acoust. Soc. Am. **90**, 1797–1805.
Sekiyama, K., and Tohkura, Y. (**1993**). "Inter-language differences in the influence of visual cues in speech perception," J. Agric. Sci. **21**, 427–444.
Sekiyama, K., Tohkura, Y., and Umeda, M. (**1996**). "A few factors which affect the degree of incorporating lip-read information into speech perception," in Proceedings of the 4th International Conference on Spoken Language Processing, Philadelphia, PA, pp. 1481–1484.
Soto-Faraco, S., Navarra, J., Voloumanos, A., Sebastián-Gallés, N., Weikum, W., and Werker, J. F. (**2007**). "Discriminating languages by speechreading," Percept. Psychophys. **69**, 218–231.
Sumby, W., and Pollack, I. (**1954**) "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Summerfield, Q. (**1979**). "Use of visual information for phonetic perception," Phonetica **36**, 314–331.
Takata, Y., and Nábelek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.
Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., and Jones, C. J. (**1977**). "Effects of training on the visual recognition of consonants," J. Speech Hear. Res. **20**, 130–145.
Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., and Werker, J. F. (**2007**). "Visual language discrimination in infancy," Science **316**(5828), 1159.
Werker, J. F., Frost, P. E., and McGurk, H. (**1992**). "La langue et les levres: Cross-language influences on bimodal speech perception," Can. J. Psychol. **46**, 551–568.

Nonuniform speaker normalization using affine transformation

S. V. Bharath Kumar[a)]
Department of Electrical and Computer Engineering, University of California-San Diego, La Jolla, California 92093-0407

S. Umesh[b)]
Department of Electrical Engineering, Indian Institute of Technology, Kanpur 208016, India

(Received 6 July 2006; revised 31 May 2008; accepted 2 June 2008)

In this paper, a well-motivated nonuniform speaker normalization model that affinely relates the formant frequencies of speakers enunciating the same sound is proposed. Using the proposed affine model, the corresponding universal-warping function that is required for normalization is shown to have the same parametric form as the mel scale formula. The parameters of this universal-warping function are estimated from the vowel formant data and are shown to be close to the commonly used formula for the mel scale. This shows an interesting connection between nonuniform speaker normalization and the psychoacoustics based mel scale. In addition, the affine model fits the vowel formant data better than commonly used *ad hoc* normalization models. This work is motivated by a desire to improve the performance of speaker-independent speech recognition systems, where speaker normalization is conventionally done by assuming a linear-scaling relationship between spectra of speakers. The proposed affine relation is extended to describe the relationship between spectra of speakers enunciating the same sound. On a telephone-based connected digit recognition task, the proposed model provides improved recognition performance over the linear-scaling model. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2951597]

PACS number(s): 43.72.Ar, 43.72.Ne [DOS] Pages: 1727–1738

I. INTRODUCTION

Over the years, there has been much interest in trying to understand the fact that phonologically identical utterances show a great deal of acoustic variation and yet listeners are able to recognize words spoken by different speakers despite these variations. Many approaches have been proposed to reduce the interspeaker variation in the acoustic data especially for formants of vowels. These include the formant-ratio theory, which is based on the idea that vowels are relative patterns and not absolute formant frequencies. Some of these formulations include those by Syrdal and Gopal (1983) and Miller (1989),

$$\text{Syrdal – Gopal: } \mathcal{B}(F_1)-\mathcal{B}(F_0),\quad \mathcal{B}(F_2)-\mathcal{B}(F_1),$$

$$\mathcal{B}(F_3)-\mathcal{B}(F_2),$$

$$\text{Miller: } \log\left(\frac{F_1}{\text{SR}}\right),\quad \log\left(\frac{F_2}{F_1}\right),\quad \log\left(\frac{F_3}{F_2}\right),$$

where $\mathcal{B}(F_i)$ is the "Bark" equivalent of the ith formant frequency F_i and SR is the sensory reference derived from the geometric mean of F_0 over an interval of time. Nearey (1978) used constant log-interval normalization given by

$$\text{Nearey: } \log(F_1)-\mu_L,\quad \log(F_2)-\mu_L,\quad \log(F_3)-\mu_L,$$

where μ_L is mean log value of the speakers' F_1 and F_2.

Bladon *et al.* (1983) proposed a normalization method based on the observation that the average frequency difference between formants of vowels produced by men and those produced by women approximately differ by 1 bark. This method uses the whole spectrum and not just the formant frequencies. Therefore, their method of normalization involved shifting down the auditory spectrum produced by women by about 1 bark.

Another approach to look at the problem of vowel normalization is to use the idea of vocal-tract length normalization. Interspeaker variations are attributed to the physiological differences in the vocal tracts of the speakers. Nordström and Lindblom (1975) proposed a normalization procedure in which the formants are scaled by a constant scale factor based on the estimate of the speaker's average vocal-tract length in open vowels, as determined from measurement of F_3. This is usually referred to as uniform scaling. In this approach, all the formant frequencies of the subject to be normalized are simply divided by the factor $\alpha=\hat{F}_{3,s}/\hat{F}_{3,r}$. $\hat{F}_{3,s}$ and $\hat{F}_{3,r}$ are the *average* F_3 of open vowels of the subject s and the male reference speaker r, respectively. Fant (1975) argued that uniform scaling is a very crude approximation and proposed that the scale factor be made a function of both formant number and vowel category. With this approach, Fant (1974) claimed to reduce the female-male variance to one-half of that remaining after the simple uniform scaling based normalization suggested by Nordström and Lindblom (1975). Nearey (1978) extensively studied the validity of linear-scaling model and concluded that there may be some

[a)]This work was done while the author was a graduate student at the Department of Electrical Engineering, Indian Institute of Technology, Kanpur 208016, India. URL: http://ieng9.ucsd.edu/~bsriperu. Electronic mail: bharathsv@ucsd.edu

[b)]URL: http://home.iitk.ac.in/~sumesh. Electronic mail: sumesh@iitk.ac.in

systematic speaker-dependent variation supporting some of Fant's (1974) observations. However, his efforts to find a better additive transform than log transform (which corresponds to linear scaling) using generalized linear models did not yield any alternate scale (Nearey, 1992).

In speaker-independent automatic speech recognition (ASR), there is considerable interest in speaker normalization since one of the major factors affecting recognition performance is the acoustic variability in similar enunciations by different speakers. To improve the performance of these systems and to make them approach that of speaker-dependent systems, various speaker normalization procedures based on spectral warping (Kamm *et al.*, 1995; Lee and Rose, 1998; Wakita, 1977; Wegmann *et al.*, 1996) have been proposed in speech recognition literature. Most of these approaches linearly scale the frequency axis of spectra of the given utterance to compensate for differences in the formant positions between speakers. We refer to these normalization procedures as a linear scaling or uniform normalization since all frequencies are scaled or normalized by a *constant* scale factor. The motivation for linear scaling comes from the fact that to a first-order approximation, the vocal-tract shape can be assumed to be a tube of uniform cross section, and for this simplifying approximation, the formant frequencies are inversely proportional to the vocal-tract length (Wakita, 1977). Although we often refer to the relation between formants, in speech recognition, the same relation is extended to the entire spectral envelope since the features in ASR are usually derived from the spectral envelope.

The simple linear-scaling model discussed above neglects both the location of constrictions and the vocal-tract shape. Motivated by Fant's (1975) work, there has been considerable interest in nonlinear scaling in speech recognition literature. Since Fant's (1975) method requires the knowledge of phoneme and formant number, it cannot be directly used in ASR. Therefore, in ASR, different *ad hoc* nonlinear frequency-scaling functions have been proposed (Acero and Stern, 1991; Eide and Gish, 1996; McDonough *et al.*, 1998; Zhan and Waibel, 1997; Zhan and Westphal, 1997), where the choice of a nonlinear frequency-scaling function is mostly driven by the ease of implementation and parsimony of parameters. However, in most of these *ad hoc* schemes, there is no specific physiological or acoustic motivation for the choice of a particular parametric form.

In this paper, we propose a well-motivated nonlinear frequency-scaling function that affinely relates the formant frequencies between speakers enunciating the same sound. The motivation for the proposed model comes from our previous experiments (Umesh *et al.*, 2002b) on *actual* speech data where we empirically obtained a piecewise approximation to a warping function that separates the speaker-dependent factor as a translation factor. In this paper, we derive the mathematical form of the universal-warping function associated with the proposed affine model and show that it is also similar in shape to the empirically obtained warping function. The proposed model is compared with other linear and nonlinear models commonly used in speech recognition and is shown to provide the best fit to formant data. The efficacy of the method in speech recognition is demonstrated on a telephone data based connected digit recognition task where the affine method provides improved recognition performance than the conventional uniform scaling model. We conclude by pointing out to the similarity of universal-warping function associated with the affine model (based on speech data alone) and mel-warp function (based on hearing experiments), thereby showing an interesting connection between psychoacoustics and nonuniform speaker normalization.

II. REVIEW OF UNIVERSAL-WARPING FUNCTION AND RELATION BETWEEN NORMALIZATION SCHEMES

In this section, we review our previous work (Umesh *et al.*, 2002a, 2002b), where we have approached the problem of speaker normalization through the concept of *universal-warping* function. The basic idea of universal-warping function is to find a warping function of the frequency axis that maps the physical frequency, f, to an alternate domain, ν, such that in the alternate domain the speaker-dependent parameter separates out as a pure translation factor. It is universal in the sense that the same mapping should be applicable to *all speech data* irrespective of the speaker. Throughout this paper, when we talk of spectra (i.e., functions of frequency), we use f to denote the frequency variable, and when we talk of formants (i.e., specific frequency values), we use F_i to denote the ith formant frequency. The universal-warping approach assumes that the speaker dependencies can be modeled through a single translation factor. For example, the commonly assumed linear-scaling model where all formant frequencies (F_i) between speakers r and s are scaled by one constant α_{rs} is given by

$$F_{i,\mathrm{r}} = \alpha_{\mathrm{rs}} F_{i,\mathrm{s}}. \tag{1}$$

In this case, α_{rs} is the speaker-dependent parameter that relates speaker s with the reference speaker r. This can be equivalently expressed in the log-warped domain as

$$\log(F_{i,\mathrm{r}}) = \log(\alpha_{\mathrm{rs}}) + \log(F_{i,\mathrm{s}}), \tag{2}$$

where the speaker-dependent α_{rs} separates out as a translation factor, i.e., $\log(\alpha_{\mathrm{rs}})$. For the uniform scaling model of Eq. (1), we see from Eq. (2) that the universal-warping function is the log warping, and in this case, the speaker-dependent scale factor separates out as a translation factor in the log-warped domain. Equivalently, if we take the ratio of formants (say, ith and jth) for the same speaker, we have

$$\log\left(\frac{F_{i,\mathrm{r}}}{F_{j,\mathrm{r}}}\right) = \log(F_{i,\mathrm{r}}) - \log(F_{j,\mathrm{r}})$$

$$= \log\left(\frac{\alpha_{\mathrm{rs}} F_{i,\mathrm{s}}}{\alpha_{\mathrm{rs}} F_{j,\mathrm{s}}}\right) = \log(F_{i,\mathrm{s}}) - \log(F_{j,\mathrm{s}}). \tag{3}$$

Hence Miller's (1989)approach to normalization, which is similar to the above equation (except for F_1/SR), is equivalent to uniform or linear scaling. Nearey's approach is also a variation of the uniform scaling model, with μ_L corresponding to the speaker-dependent shift factor. We recently became aware of the work of Nearey in Nearey (1978), where he introduced the use of the log-additive hypothesis. Our

concept of universal-warping function is similar to this concept, except that we are looking in a generalized framework of translation for any transformation and not necessarily log transformation. Note that since the shift factor does not depend on any specific phoneme, and only on the speaker, it is often referred to as *extrinsic* normalization factor. Extending the model in Eq. (1) to spectral envelopes, we can assume that the spectral envelopes of r and s are scaled versions of one another, i.e., $P_r(f)=P_s(\alpha_{rs}f)$. In the case of spectral envelopes, log-warping the frequency-axis, i.e., $\lambda=\log(f)$ results in

$$\mathbb{P}_r(\lambda) = P_r(e^{\lambda}) = P_s(e^{\lambda+\log\,\alpha_{rs}}) = \mathbb{P}_s(\lambda + \log\,\alpha_{rs}), \tag{4}$$

where $\mathbb{P}_r$ and $\mathbb{P}_s$ are log-warped versions of P_r and P_s, respectively. Therefore, the frequency-warped spectral envelopes are *shifted* versions of each other if the model in Eq. (1) is indeed true.

From experiments of Fant (1975) and some of our previous experiments in Umesh *et al.* (2002a, 2002b, 2002c), it has been observed that there are deviations from the uniform scaling model. Since there are deviations from the uniform scaling model, log warping is *not* the appropriate universal-warping function to separate the speaker-dependent parameter. In Umesh *et al.* (2002b, 2002c), a piecewise approximation to the universal-warping function was found empirically from speech data, such that in the universal-warped domain, the same sound enunciated by different speakers were translated versions of one another. This empirically obtained universal-warping function is referred to as the *speech scale*. Interestingly, the speech scale was found to be "very similar" to the mel scale.

Note that since Bark scale is similar to mel scale (and hence speech scale), the normalization method of Bladon *et al.* (1983) also uses a similar idea, which involves shifting down the auditory spectrum produced by women by about 1 bark. While Bladon *et al.* (1983) used a gender-specific shift, every speaker has a speaker-specific relative shift in the speech scale in the approach of Umesh *et al.* (2002b, 2002c). Further, this normalization approach is very similar to that proposed by Syrdal and Gopal (1983). Note that the empirical speech scale, the mel scale, and the Bark scale are very similar, and hence, we have the following normalization

$$\mathcal{S}(F_{i,r}) - \mathcal{S}(F_{i,s}) \approx \mathrm{mel}(F_{i,r}) - \mathrm{mel}(F_{i,s}) = c_{rs}, \tag{5}$$

where $S(.)$ is the speech scale and c_{rs} is a speaker-dependent constant that is independent of i, i.e., formant number.

In this paper, our goal is to find the relation between spectra of the same sound enunciated by different speakers, such that the corresponding universal-warping function is similar to our empirically obtained speech scale.

III. PROPOSED AFFINE TRANSFORMATION: RELATION BETWEEN FORMANTS OF SPEAKERS

We propose the following affine model to describe the relation between the formant frequencies of subject and reference speakers, i.e.,

$$(F_r + \kappa) = \alpha_{rs}(F_s + \kappa), \tag{6}$$

where F_r and F_s are the formant frequencies of the reference speaker r and subject speaker s, respectively. α_{rs} is a speaker-dependent parameter relating the reference and subject speakers. κ is a constant in the model and is *not* dependent on the speakers. Equivalently, we can also rewrite Eq. (6) as

$$F_r = \alpha_{rs}F_s + \kappa(\alpha_{rs} - 1). \tag{7}$$

We can clearly see the affine relation in the above equation, and hence we refer to the proposed model as the affine model. Note that unlike the conventional affine equation the shift factor is *also* a function of the scaling factor α_{rs}. As before, the same model is also used to describe the relationship between spectral envelopes. Before studying the model, let us first discuss the motivation behind its proposal.

A. Motivation

The proposed affine model relating the formant frequencies of speakers enunciating the same sound is broadly motivated by the following observations.

(1) We are interested in finding a parametric model relating the formants of speakers such that the corresponding universal-warping function has a parametric form similar to the mathematical formula for the mel scale since the mel scale closely matches the empirically obtained piecewise linear speech scale.
(2) It is well known that the higher-order formants are mostly affected by the vocal-tract length and so the first-order approximation of linear scaling is valid at higher frequencies. Therefore, the choice of a parametric model should capture this behavior at higher frequencies where the scale factor should almost be constant between any two speakers. From Eq. (6), we see that this is approximately true for F_r, $F_s \gg \kappa$, in which case $F_r \approx \alpha_{rs}F_s$.
(3) Finally, at low frequencies (lower formants), observation on Texas Instruments-Massachusetts Institute of Technology (TIMIT) data (Umesh *et al.*, 2002a) has shown that there are significant deviations from uniform scaling with the trend being increasing dilation/compression with decreasing frequencies (lower formants). A similar trend in the values of warp factors at low frequencies has been noticed by Potamianos and Narayanan (2003). Nearey also observes a weak indication of a roughly sinusoidal pattern of deviation for low frequency formants in Sec. IV C of Nearey (1978). As we will show later, the proposed model reflects the above trends also.

For the proposed affine model in Eq. (6), analogous to uniform scaling, our goal is to find a one-to-one transformation, $\nu=\Phi(f)$ such that

$$\mathbb{P}_r(\nu) = P_r[\Phi^{-1}(\nu)]$$
$$= P_s\{g[\alpha_{rs}, \Phi^{-1}(\nu)]\Phi^{-1}(\nu)\} = \mathbb{P}_s(\nu + \varsigma_{rs}), \tag{8}$$

where $g[\cdot]$ is both a frequency- and a speaker-dependent scaling function. ς_{rs} is only speaker dependent and is independent of frequency. In this case, $\nu=\Phi(f)$ is the universal-warping function for the affine model.

B. Universal frequency-warping function for affine model

The affine model that we have proposed in this paper in Eq. (6) can be rewritten as

$$\log\left(1+\frac{F_r}{\kappa}\right)=\log\,\alpha_{rs}+\log\left(1+\frac{F_s}{\kappa}\right),\quad \kappa\neq 0. \tag{9}$$

With $\nu=\log(1+f/\kappa)$, we have $\nu_r=\nu_s+\log\,\alpha_{rs}$, where ν_r and ν_s are the warped frequencies of $f=F_r$ and $f=F_s$, respectively. Hence, the warped frequencies appear as shifted or translated versions in the ν-domain and the translation factor is speaker dependent. Therefore, the universal-warping function for the affine model is

$$\nu=\Phi(f)=\log\left(1+\frac{f}{\kappa}\right). \tag{10}$$

We are interested in the relation between $\Phi(f)$ and the mel scale [which from Umesh *et al.* (2002a) we know is close to the empirical speech scale].

Stevens and Volkman (1940) experimentally obtained a nonlinear mapping between perceived and physical frequencies of a tone and referred to it as the mel scale. In their original work, Stevens and Volkman (1940) had experimentally obtained the mapping at discrete set of points, for which various closed-form curves have been fitted by researchers. The widely accepted closed-form approximations to mel scale have the functional form

$$\eta=a\,\log_{10}\left(1+\frac{f}{b}\right), \tag{11}$$

where f is in Hz and η is in *mels*. Fant's technical mel formula is defined with $a=1000/\log 2$ and $b=1000$, whereas in speech recognition the widely used formula is defined with $a=2595$ and $b=700$, i.e.,

$$\eta_{\text{mel}}=\Theta(f)=2595\,\log_{10}\left(1+\frac{f}{700}\right). \tag{12}$$

Although in speech recognition, the mel scale is the most commonly used psychoacoustic scale, in many other areas of speech the bark scale or equivalent rectangular bandwidth (ERB) scale is usually used. Note that these two scales also have a functional form similar to Eq. (11). Therefore, for the purposes of this paper, we will only compare with the mel scale. We point out that Nearey (1978) also used a transformation of the type $\log(f+b)$, where b is a constant which is equivalent to our proposed universal-warp function given in Eq. (10). The warping function, $\Phi(\cdot)$ obtained from the affine model and the mel-warp function, $\Theta(\cdot)$ have a similar parametric form, but their behavioral similarity depends on the value of κ (or, equivalently, of b). It can be seen that $\nu\rightarrow\eta_{\text{mel}}/2595\,\log_{10}e$ for "large" values of κ and $\nu\rightarrow\lambda=\log(f)$ for "small" values of κ.

As mentioned previously, we can extend the relationship for formants in Eq. (9) to spectral envelopes also. We can rewrite the affine model as, $f'=\alpha_{rs}f+\kappa(\alpha_{rs}-1)$. It is easy to see that in the warped domain, $\nu=\Phi(f)=\log(1+f/\kappa)$, the spectral envelopes are shifted versions of each other, i.e.,

$$\begin{aligned}\mathbb{P}_r(\nu)&=P_r[f=\kappa(e^{\nu}-1)]\\&=P_s[f'=\alpha_{rs}f+\kappa(\alpha_{rs}-1)]\\&=P_s[\kappa\alpha_{rs}(e^{\nu}-1)+\kappa(\alpha_{rs}-1)]\\&=P_s[\kappa(e^{\log\,\alpha_{rs}+\nu}-1)]=\mathbb{P}_s(\nu+\log\,\alpha_{rs}).\end{aligned} \tag{13}$$

In Eqs. (10) and (13), we assume that κ is a constant and also that it is a speaker-independent parameter. However, the value of this constant κ is yet to be determined. In the next subsection, we describe in detail how we have estimated the value of κ using actual formant data.

C. Estimation of κ and α

The estimation of κ is carried out using the formant data from Peterson and Barney (1952) (PnB) and Hillenbrand *et al.* (1995) (HiL) vowel databases. The PnB database consists of 76 speakers (33 males, 28 females, and 15 children), with each speaker contributing two utterances for each of ten vowels (/aa/, /ae/, /ah/, /ao/, /eh/, /er/, /ih/, /iy/, /uh/, /uw/). Alternately, we can consider the PnB database as having 152 speakers (66 males, 56 females, and 30 children), with each speaker uttering ten vowels once. The HiL database *effectively* consists of 98 speakers (37 males, 33 females, 13 boys, and 15 girls), with each of them speaking only once for each of 12 vowels (/ae/, /ah/, /aw/, /eh/, /ei/, /er/, /ih/, /iy/, /oa/, /oo/, /uh/, /uw/). We have not considered other speakers in the HiL database since some of their formant estimates have been marked zero.

Every speaker in both of these databases is characterized by formant frequencies (F_1,F_2,F_3) for each vowel and we create a "formant vector," $\mathbf{f}$, by concatenating the formant frequencies of different vowels spoken by that speaker. Therefore, each speaker is represented by a 30-dimensional formant vector for PnB (since there are ten vowels and three formats for each vowel) and by a 36-dimensional formant vector for HiL, which has 12 vowels. Let $\mathbf{f}_{s,j}$ be the formant vector of the jth subject speaker of a given database. Similarly, let $\mathbf{f}_r$ be the formant vector of the reference speaker. (We will discuss the selection of the reference speaker shortly.) $\mathbf{f}_r$ and $\mathbf{f}_{s,j}$ are defined as

$$\mathbf{f}_r=[F_{r,1}\quad F_{r,2}\quad\ldots\quad F_{r,n}]^T,$$

$$\mathbf{f}_{s,j}=[F_{s,j1}\quad F_{s,j2}\quad\ldots\quad F_{s,jn}]^T,\quad j=1,2,\ldots,M,$$

where $n=30$ and $M=152$ for PnB, and $n=36$ and $M=98$ for HiL. We have adopted the following approach to estimate the value of κ. Initially we assume κ to be speaker dependent. From Eq. (7), the predicted formant frequency vector of the reference with respect to the jth subject speaker, i.e., $\hat{\mathbf{f}}_{r,j}$, is

$$\hat{\mathbf{f}}_{r,j}=\alpha_j\mathbf{f}_{s,j}+\kappa_j(\alpha_j-1)\mathbf{1}, \tag{14}$$

where α_j and κ_j are the parameters of the jth speaker and $\mathbf{1}$ is the n-dimensional vector $[\underset{n}{1\ 1\ \ldots 1}]^T$. Note that κ is also a function of j, the subject speaker; i.e., initially we assume κ to be speaker dependent. The cost function to be minimized is

$$\|\mathbf{f}_r - \hat{\mathbf{f}}_{r,j}\|^2 = \sum_{i=1}^{n} [F_{r,i} - \alpha_j F_{s,ji} - \kappa_j(\alpha_j - 1)]^2, \quad (15)$$

which is quadratic in α_j and κ_j. α_j and κ_j are estimated by minimizing Eq. (15) over $\mathbb{S}=\{(\alpha,\kappa)\in\mathbb{R}^2_{++}\}$ for each subject speaker, i.e.,

$$(\alpha_j, \kappa_j) = \arg\min_{\mathbb{S}} \|\mathbf{f}_r - \hat{\mathbf{f}}_{r,j}\|^2. \quad (16)$$

As one reviewer suggested, we could have tried a joint optimization of α and a fixed κ over all speaker pairs, but there is no easy way to set non-negativity constraints for α and κ. These constraints come from our empirical observations (discussed below) that κ takes mostly positive values and α lies in the conventional range of 0.7–1.3. Another reviewer had suggested the use of generalized linear models and a variance stabilizing transform and had some doubts about the suitability of our approach for κ estimation. One of the main reasons for the approach that we have followed below is due to the fact that the variance of κ estimates are affected by the value of the corresponding α in the affine model. We describe the approach below.

In Eqs. (14) and (15), we see that for the case of $\alpha=1$, κ can take *any* value without affecting the minimization. In general, the variance of κ becomes extremely large as $\alpha \to 1$. Therefore, to get reliable estimates of κ, we should consider subject speakers and reference speakers such that their corresponding α are different from unity. Note that this constraint is necessary *only* to get reliable estimates of κ. Once, the speaker-independent κ parameter has been estimated, we can take any pair of speakers and their corresponding α can be unity. We have taken male-female and male-child combinations of speakers for the estimation of κ. Figure 1(a) shows the estimates of κ for the 37 male speakers of the Hillenbrand database with *each* of the 33 female speakers acting as a reference speaker. Similarly, Fig. 1(b) shows the estimates of κ for the male speakers, with each child speaker acting as a reference. Figures 2(a) and 2(b) show the corresponding α estimates. In most of these cases, α are quite different from unity, and therefore the variance of κ are reasonably small. Further, as expected, we see that there is a higher variance for κ estimates when females are used as reference when compared to child reference speakers since the α values are closer to unity. We also see in Fig. 1 that the κ values are positive for most speaker pairs.

In the above experiment, we have taken each female (or child) speaker in the database as a reference speaker. For each subject speaker, we can average the estimates of α and κ over the set of all female speakers, which would correspond to using an average female speaker. Instead of taking the sample mean, we use the following procedure. Let there be M (male) subject speakers and K female speakers. Let α_{ij} and κ_{ij} be the parameters of the jth subject with respect to the ith female speaker as reference. From Eq. (6), we have

$$\hat{\mathbf{f}}_{r,ij} = \alpha_{ij}\mathbf{f}_{s,j} + \kappa_{ij}(\alpha_{ij} - 1)\mathbf{1}, \quad (17)$$

where $i=1,2,\ldots,K$ and $j=1,2,\ldots,M$. $\hat{\mathbf{f}}_{r,ij}$ is the predicted formant frequency vector of the ith female speaker with respect to the jth subject speaker. α_{ij} and κ_{ij} are estimated as $(\alpha_{ij},\kappa_{ij})=\arg\min_{\mathbb{S}}\|\mathbf{f}_{r,i}-\hat{\mathbf{f}}_{r,ij}\|^2$. The predicted formant frequency vector of the average female (reference) speaker with respect to the jth subject speaker is given by

$$\hat{\mathbf{f}}_{r,j} = \frac{1}{K}\sum_{i=1}^{K}\hat{\mathbf{f}}_{r,ij} = \frac{1}{K}\sum_{i=1}^{K}[\alpha_{ij}\mathbf{f}_{s,j} + \kappa_{ij}(\alpha_{ij}-1)\mathbf{1}]$$
$$= \left(\sum_{i=1}^{K}\frac{\alpha_{ij}}{K}\right)\mathbf{f}_{s,j} + \left(\sum_{i=1}^{K}\frac{\kappa_{ij}(\alpha_{ij}-1)}{K}\right)\mathbf{1}. \quad (18)$$

It is clear from Eqs. (14) and (18) that

$$\alpha_j = \frac{1}{K}\sum_{i=1}^{K}\alpha_{ij} \text{ and } \kappa_j = \frac{\Sigma_{i=1}^{K}\kappa_{ij}(\alpha_{ij}-1)}{\Sigma_{i=1}^{K}(\alpha_{ij}-1)}. \quad (19)$$

By this approach, the estimate of κ_j has considerably less variance. The value κ_j of each male speaker with respect to the average female (reference) speaker is shown in Fig. 3(a). Similarly, the κ_j of each male speaker with respect to the average child speaker is shown in Fig. 3(b). We can find the average κ for the database by averaging over all M subject speakers, i.e., $\kappa_{mean}=1/M\Sigma_{j=1}^{M}\kappa_j$. The unconstrained mean value of κ (for the Hillenbrand database) using a male-female combination is 575 and using a male-child combina-

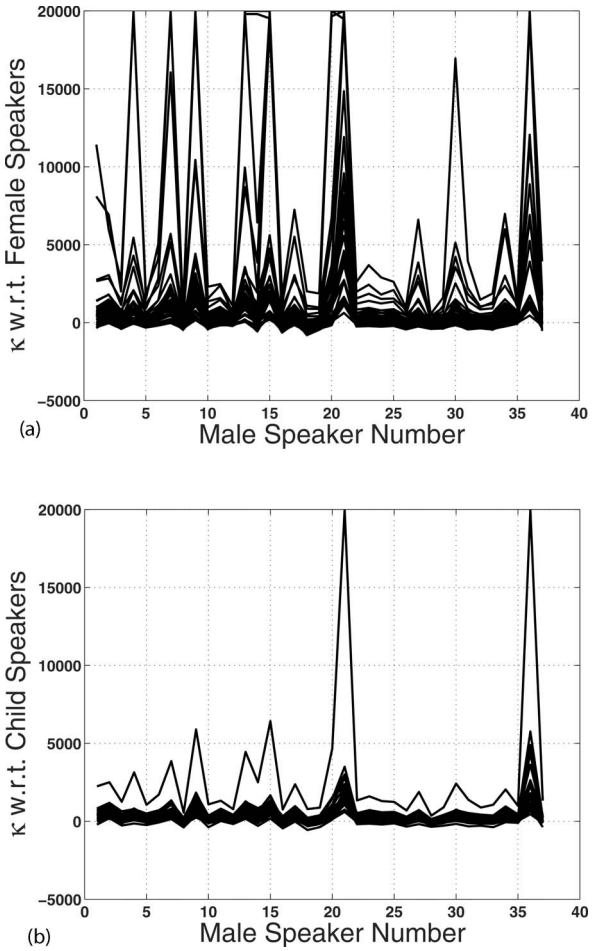


FIG. 1. Estimates of κ obtained for the 37 male speakers in Hillenbrand data with each of the (a) female speakers and (b) child speakers in the database as reference. No constraint has been put on the range of κ and was searched over a very wide range of $[-20000, +20000]$.

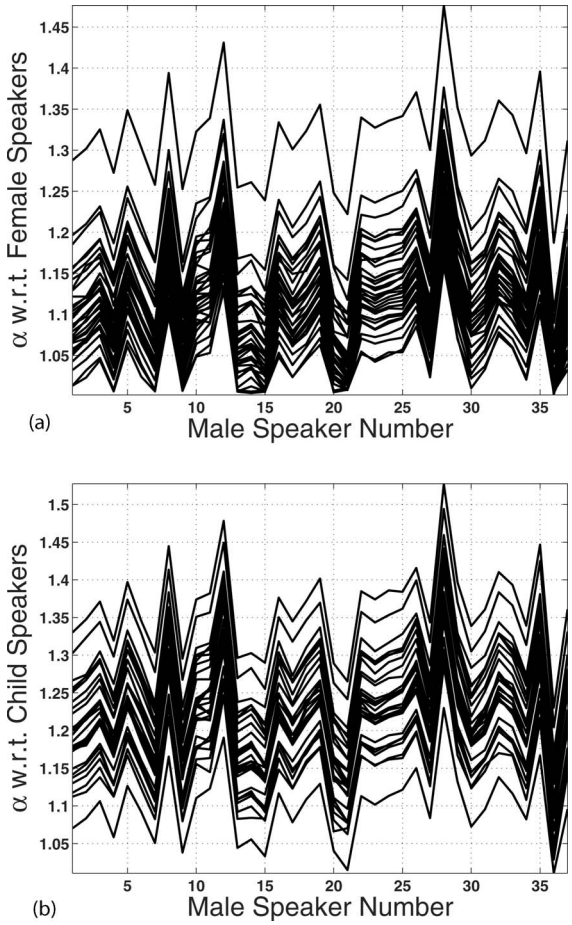


FIG. 2. Estimates of α obtained for the 37 male speakers in Hillenbrand data with each of the (a) female speakers and (b) child speakers in the database as reference. No constraint has been put on the range of α and was searched over a very wide range of $[-20000, +20000]$.

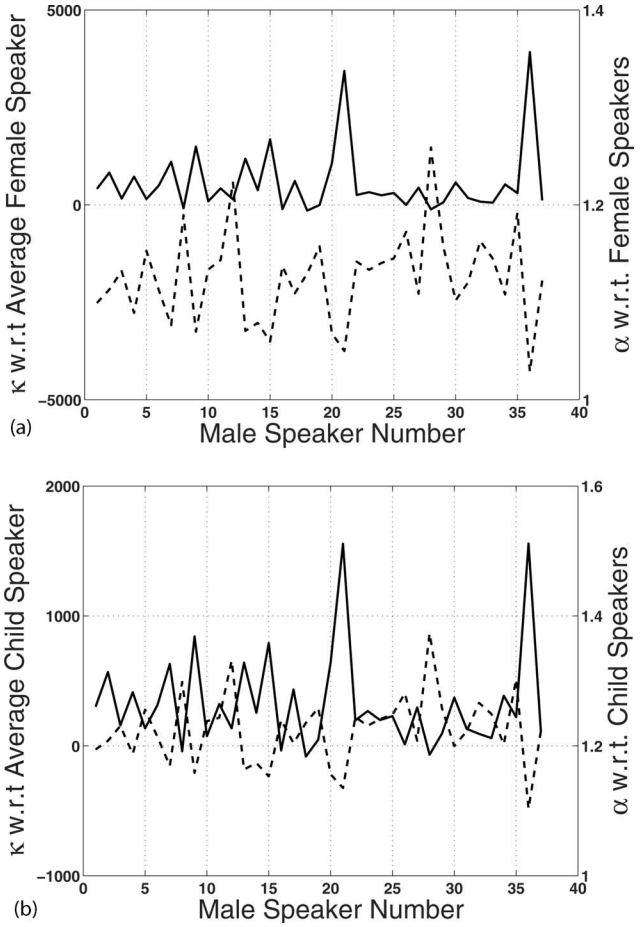


FIG. 3. Estimates of κ and α obtained for the 37 male speakers in Hillenbrand data with (a) the average female speaker and (b) the average child speaker in the database as reference. No constraint has been put on the range of α and κ and were searched over a very wide range of $[-20000, +20000]$.

tion is 330. If one leaves out the two outliers in the male-female experiment, then the mean κ is 370, which is in the same range as that obtained using a male-child combination.

Alternately, we have also computed the mean value of κ using *all* (instead of only male) speakers from a given database with respect to the female (or child) speakers as reference. Note that in this case, it is possible that for some pairs of speakers the α values are close to unity, with the corresponding κ estimates having high variance. Therefore, we chose only those pairs of speakers whose κ estimates were in the interval between 0 and 2000. This is reasonable since from Figs. 3(a) and 3(b), we see that reliable estimates of κ lie mostly in the region of 0–2000. Once again, we used individual female (or child) speakers as reference speakers and all the 98 speakers in the database as subject speakers. We then obtain averaged values of α_j and κ_j over all the female (or child) reference speakers using Eq. (19). Having estimated κ_j, its mean estimate, κ_{mean} is computed for a given database. The value of κ_{mean} has been computed to be 508.04 for the PnB database and 495.67 for the HiL database. Similarly the κ_{mean} for child reference speakers was found to be 434.18 for the PnB database and 548.28 for the HiL database. As we will show both in terms of warping function behavior and speech recognition performance, these small variations in κ do not have a significant effect.

The warping functions ν_{PnB} and ν_{HiL} for PnB and HiL data, respectively, are given by

$$\nu_{\text{PnB}} = \log\left(1 + \frac{f}{508.04}\right), \tag{20a}$$

$$\nu_{\text{HiL}} = \log\left(1 + \frac{f}{495.67}\right). \tag{20b}$$

As we mentioned previously, Nearey (1978) also investigated the use of a transformation of the type $\log(f+b)$, which is similar to our affine model. The main difference between our approach and Nearey (1978) is that the latter considers only the average formant values for males, females, and children for several languages including the American English database of Peterson and Barney (PnB). On the other hand, we have considered *pairwise* all the speakers in the PnB and Hillenbrand (HiL) databases. We have then averaged the estimates using Eq. (19). Further, unlike Nearey (1978), we have considered all the formants and not just the first or second formant. One of the reasons Nearey was motivated to use the transform of the type $\log(f+b)$ was the systematic speaker-dependent variation in acoustic parameters. However, his analysis using first and second formants in Nearey (1978, 1992) did not show any

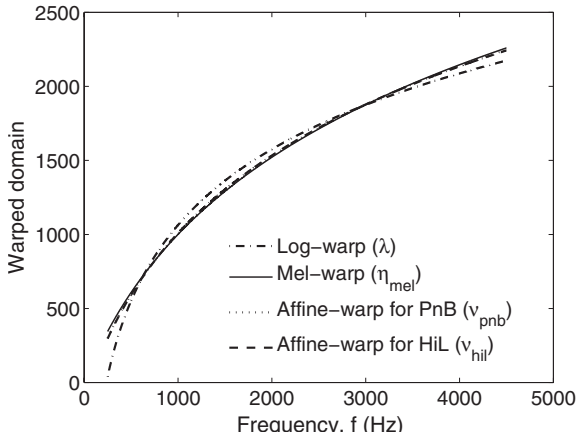


FIG. 4. Comparison of different universal-warping functions including the affine-warp, the log-warp, and the mel-warp functions. The affine-warp functions for Peterson and Barney (PnB) and Hillenbrand (HiL) databases almost overlap since they are functionally similar, as seen from Eqs. (20a) and (20b).

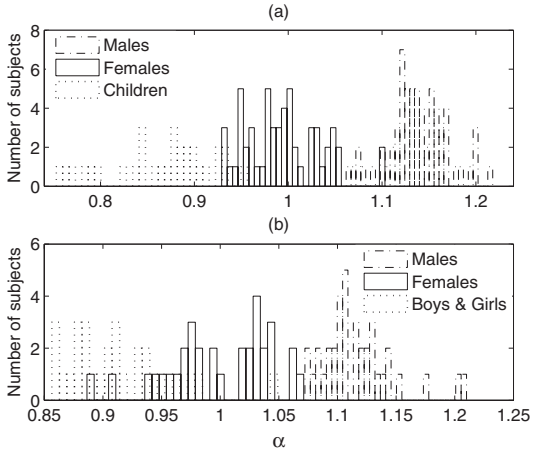


FIG. 5. Histogram of the speaker-dependent parameter α in speaker normalization using the affine model for (a) the Peterson and Barney database and (b) the Hillenbrand database.

substantial improvement by using other frequency scales over the log scale. In this paper, however, we show that there is an improvement in speech recognition performance using our proposed affine model when compared to the linear-scaling model.

Equations (20a) and (20b) shows that the universal-warping functions obtained for the affine model from PnB and HiL databases are essentially the same and are functionally closer to the mel-warp function than to the log-warp function. Figure 4 shows the plot of log-warp (λ), mel-warp (η_{mel}), and universal-warp ($\nu_{\text{PnB}}, \nu_{\text{HiL}}$) functions for the affine model. Since the value of κ is almost the same for both PnB and HiL databases, the universal-warp function for both of these databases appear the same. It is interesting to note that the affine-warp functions are almost the same as the mel-warp function that is obtained by fitting a curve to psychoacoustic experimental data of Stevens and Volkman (1940) whose equation is given in Eq. (12). Therefore, our experiment on nonuniform speaker normalization, which is conducted on speech data alone, shows the required frequency warping to be close to the mel scale. This study thus justifies the use of mel scale in speech recognition, not only from the psychoacoustic point of view but also from the viewpoint of nonuniform speaker normalization. We compare the normalization performance of affine-warp functions with respect to mel warp and log-warp functions in terms of word error rate in Sec. IV.

Apart from the nature of warping function, we are also interested in studying the distribution of reestimated α values across genders. Using the fixed value of κ_{mean} (which is assumed to be speaker independent), we reestimated α for each speaker. Figure 5 shows the histograms of α for male, female, and child speakers of PnB and HiL databases. The trend in the estimates of α across genders shows the existence of *gender separability*. Also, since the average female is considered as the reference subject, the female speakers are centered around an $\alpha=1$ warping factor, while male and child subjects have $\alpha>1$ and $\alpha<1$, respectively. Although not shown here, a similar distribution of α estimates occurs when we use the male reference speaker. As an illustration of the normalization process, in Fig. 6, we show the smoothed spectra of a male-speaker appropriately normalized to match the female reference using the affine model. In the next subsection, we will compare the nature of frequency warping of the proposed affine model with other models commonly used in speech recognition.

D. Normalization models used in speech recognition

As we have mentioned previously, Fant's work (Fant, 1975) shows that there are significant deviations from the uniform scaling model of Eq. (1). He suggests that the scale factor be made a function of formant number and vowel category. Motivated by Fant's work, in speech recognition literature, there has been an attempt to make the scaling a function of both frequency and a speaker-dependent factor α_{rs}, i.e., $g(\alpha_{\text{rs}},f)$. So, the relationship between spectral envelopes of the reference and subject speakers is assumed to be

$$P_{\text{r}}(f) = P_{\text{s}}[g(\alpha_{\text{rs}},f)f]. \tag{21}$$

Note that we assume that there is only one speaker-dependent parameter α_{rs} for each speaker, and this factor is

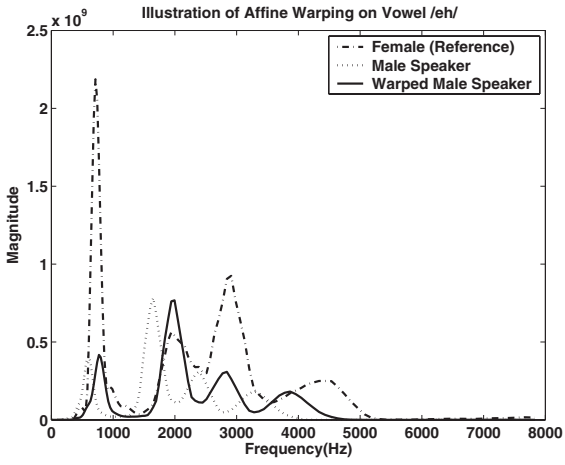


FIG. 6. Illustration of the normalization of the spectral envelope of a male-subject speaker to a female-reference speaker using the affine model.

TABLE I. Frequency-dependent scale factor $g(\alpha_{rs},f)$ between the average male and the average female for various speaker normalization schemes using the vowel formant data from Peterson and Barney (PnB) and Hillenbrand (HiL) databases. The α_{rs} are estimated using least-squares criterion.

Method	Scale factor, $g(\alpha_{rs},f)$	PnB	HiL
Uniform	$g_u(f)=\alpha_{rs}$	1.17	1.15
Affine	$g_a(f)=\alpha_{rs}+\frac{\kappa(\alpha_{rs}-1)}{f}$	1.14	1.12
Eide–Gish	$g_e(f)=\alpha_{rs}^{3f/F_N}$	1.20	1.16
Power	$g_p(f)=\left(\frac{f}{F_N}\right)^{\alpha_{rs}-1}$	0.89	0.90
Bilinear	$g_b(f)=\frac{F_N}{2\pi f}\tan^{-1}\frac{(1-\alpha_{rs}^2)\sin\left(\frac{2\pi f}{F_N}\right)}{(1+\alpha_{rs}^2)\cos\left(\frac{2\pi f}{F_N}\right)-2\alpha_{rs}}$	0.14	0.125

the same for all phonemes. Some of the scaling functions include those proposed by Eide and Gish (1996) and power and bilinear transformation (Acero and Stern, 1991; McDonough *et al.*, 1998). These are *ad hoc* models, in that they have no physiological or acoustic motivation. The main motivating factors for these models have been the ease of implementation and parsimony of parameters. Therefore, until now no study has been done to understand the nature of frequency scaling in these models. In this subsection, we compare the nature of frequency scaling of these models with the proposed affine model.

Let $g_u(\alpha_{rs},f)$, $g_a(\alpha_{rs},f)$, $g_e(\alpha_{rs},f)$, $g_p(\alpha_{rs},f)$, and $g_b(\alpha_{rs},f)$ be the frequency-dependent scale factors for uniform, affine, Eide–Gish, power, and bilinear transformation models, respectively. The mathematical form of these nonlinear frequency-scaling functions is given in Table I. F_N in these equations is the Nyquist frequency, which is 8 kHz for telephone speech.

We show the nature of the frequency-dependent scaling function for different models using the average female as the reference speaker and the average male as the subject speaker. In this subsection, we use formant data, and the average female formant vector is obtained by averaging the formant vectors of all female speakers in the database. This corresponds to the reference vector f_r. Similarly, the average male subject vector, f_s, is obtained by averaging all the male formant vectors in the database. The parameter α_{rs} for each of these scaling functions is estimated by fitting $f_r = g(\alpha_{rs},f_s)f_s$ in the least-squares sense.

Table I shows the least-squares estimate of α_{rs} for each of nonuniform scaling functions using PnB and HiL data. Figure 7 shows the plot of $g(\alpha_{rs},f)$ based on α_{rs} estimates of Table I for PnB and HiL databases. At low frequencies, many of the models (except Eide–Gish) show deviation from uniform scaling, with the nature of deviation being approximately similar to empirical observations in Umesh *et al.* (2002b) and Potamianos and Narayanan (2003). However, at higher frequencies, it is clear from Fig. 7 that only the affine model behaves similar to uniform scaling, whereas the other models exhibit completely different behaviors. Uniform scaling is more reasonable at higher frequencies since higher formants are mostly affected by the length of the vocal tract. This study again confirms that the affine model is a better model for speaker normalization when compared to other *ad hoc* normalization models and may also explain the limited success of these *ad hoc* models in terms of recognition performance when compared to uniform scaling.

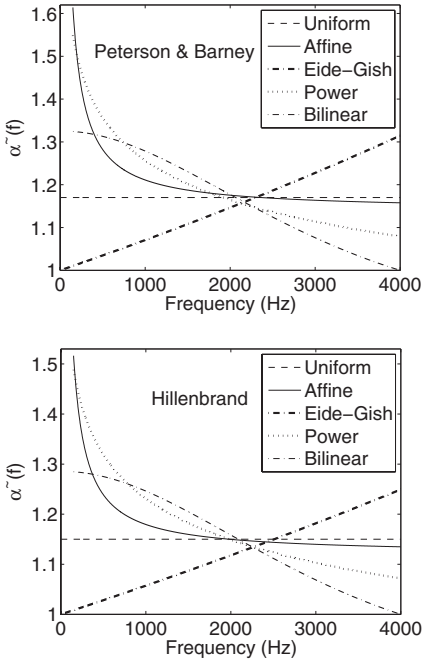


FIG. 7. Frequency-dependent scale factor $g(\alpha_{rs},f_s)$ as a function of subject formant frequency, f_s, for various speaker normalization schemes using Peterson and Barney and Hillenbrand vowel formant data (see Table I).

E. Comparison to piecewise constant scale factor model

In this subsection, we compare the universal-warping function obtained for the affine model with the empirical piecewise approximation of universal-warping function estimated in Umesh *et al.* (2002b). For the affine model, the universal-warping function is given by Eqs. (20a) and (20b) when estimated using PnB and HiL databases, respectively.

The following model is used in Umesh *et al.* (2002b) to describe the relationship between formant frequencies of two speakers enunciating the same sound,

$$P_r(f) = P_s(\alpha_{rs}^{\beta_i} f), \quad L_i \leq f \leq U_i, \tag{22}$$

where a_{rs} depends on the pair of speakers r and s and β_i is purely a function of frequency (more precisely on the chosen frequency band, i). The frequency region of interest was divided into five logarithmically equispaced bands, where L_i and U_i are lower and upper frequency boundaries of the ith frequency band. The above model can be thought of as modeling the deviation from uniform scaling by allowing β_i the freedom to change from band to band. Note that if $\beta_i=1$ for all i (or more generally if β_i is the same for all i), then one gets back the uniform scaling model of Eq. (1). It has to be noted that for the model in Eq. (22), there is no exact closed-form relationship between formant frequencies unlike the af-

fine model, where the mathematical relation is given by Eq. (6). Further, the affine model has only two free parameters (α and κ), while the piecewise approximation with five frequency bands has six parameters (β_i, $i=1,2,\ldots,5$ and a_{rs}), giving it more degrees of freedom.

In Umesh *et al.* (2002b), β_i values are estimated using TIMIT vowel data, and the corresponding piecewise approximation to universal-warping function was obtained. We refer to this scale as the speech scale. Since in this paper all the experiments have been done using PnB and HiL formant data, we have estimated the empirical universal-warping function for PnB and HiL using the procedure detailed in Umesh *et al.* (2002b). The empirically obtained curve is fitted to the parametric form in Eq. (11) using TABLECURVE2D. With a fitting accuracy of 97% for PnB and 99.6% for HiL, we obtain a continuous warping function $\Omega(f)$. The universal-warping functions, $\Omega(f)$ for PnB and HiL databases using the empirical method of Umesh *et al.* (2002b) are given as

$$\tau_{\mathrm{PnB}} = \Omega^{\mathrm{PnB}}(f) = 2364.85 \log\left(1 + \frac{f}{588.05}\right), \tag{23a}$$

$$\tau_{\mathrm{HiL}} = \Omega^{\mathrm{HiL}}(f) = 2478.24 \log\left(1 + \frac{f}{641.94}\right). \tag{23b}$$

It is interesting to note from Eqs. (23a) and (23b) that the warping function $\Omega(f)$ is behaviorally closer to affine-warp and mel-warp functions than to the log-warp function. The fact that we again get a "mel-like" curve is interesting since we have used a piecewise model to fit the formant data, which is quite different from the affine model. Since we obtain mel-like warping functions from two entirely different methods, we argue that mel scale may be important in speaker normalization.

To summarize, the affine model provides a universal-warping function that is similar to the mel curve and speech scale. In addition, at higher frequencies, the affine model behaves similar to uniform scaling, which is a desired property. It fits the formant data better than other standard nonuniform models. We therefore claim that the proposed affine model is a more appropriate model for nonuniform speaker normalization.

IV. NORMALIZATION PERFORMANCE IN ASR

In this section, we compare the normalization performance of the affine model of Eq. (6) with the commonly used linear model of Eq. (1) on a digit recognition task. We do not consider other models like Eide–Gish, power, and bilinear transformation in this comparison because of their anomalous behavior, as seen in Fig. 7. We use word error rate of the recognition system as a measure of normalization performance and study which of these models is best suited for speech recognition.

A. Task and database

The normalization experiments are evaluated on a telephone-based connected digit recognition task. The speech data for training the recognizer is derived from the *Numbers* *v*1.0*cd* corpus of Oregon Graduate Institute. The training set consists of 6078 utterances from adult male and female speakers. The performance of different normalization procedures are evaluated on two different test sets. The test set, referred to as "adults," is derived from *Numbers* corpus and consists of 2169 utterances from 790 adult male and 1379 female speakers other than those in training. The mismatched test set, referred to as "children," is derived from other than *Numbers* corpus, which is not publicly available. It consists of 2798 utterances from speakers having ages between 6 and 18 years. The children test set consists of 1225 boys and 1554 girls. All the utterances have variable digit string lengths and correspond to continuously spoken speech.

B. Front-end signal processing

Until now we have compared the models using formant data from PnB and HiL. In speech recognition the features for recognition are derived from the smoothed spectral envelope. The cepstral coefficients are the most commonly used features in state-of-the-art automatic speech recognition systems. Most of these systems use mel-frequency cepstral coefficients (MFCCs), which are obtained from mel-filter bank smoothed spectra.

In our implementation, the spectral smoothing is done using *weighted overlap segment averaging* (WOSA) (Nuttall and Carter, 1982), which is a variation of the *averaged periodogram* spectral estimation method. The details of the procedure are as follows. Speech signals are sectioned into 20 ms long frames (corresponds to 160 samples for 8 kHz sampling) with an overlap of 10 ms. A first-order backward difference signal is computed with a pre-emphasis factor of 0.97. A given pre-emphasized speech frame is segmented into a number of overlapping subframes of 64 sample width and 45 sample overlap, with each subframe being Hamming windowed. This choice of parameters is for telephone speech sampled at 8 kHz. An average autocorrelation estimate for a given frame is obtained by averaging over the autocorrelation estimates of the subframes. The Fourier transform of the averaged autocorrelation estimate is essentially the smoothed spectral envelope. This method effectively suppresses the pitch since the duration of each subframe is less than the expected pitch interval of an average adult male. In Umesh *et al.* (2004) and Sinha and Umesh (2008), we have compared this approach to conventional MFCC for vocal-tract length normalized (VTLN) and have shown slightly better performance. We have tested and found that this smoothing works well for female and children also.

In this paper, we approach the problem of speaker normalization through the concept of universal-warping function. In the universal-warped domain the speaker-dependent factor will separate out as a translation factor. Normalization is done by an appropriate shift of the universal-warped spectra. To compute the un-normalized features, the spectral features in the universal-warped domain are obtained by taking nonuniform discrete Fourier transform over the averaged autocorrelation estimate obtained from WOSA. The choice of the number of points in the warped domain is critical as it controls the resolution of spectral shift. In our implementa-

tion, 64 points spaced nonuniformly *depending on the universal-warping function* are considered between 270 and 3850 Hz. The universal-warped spectrum is then log compressed (in magnitude), and finally a discrete cosine transform (DCT) is applied to obtain cepstral features.

We now discuss the computation of features during normalization. In conventional speaker normalization, the linear-scaling model of Eq. (1) is used. To obtain the normalized features, the speech signal is warped by resampling (Kamm *et al.*, 1995) and then the mel-filter bank smoothing is applied to get VTLN features. This is more efficiently implemented by appropriately VTLN-warping the mel-filter bank (Lee and Rose, 1998).

In our proposed approach to speaker normalization, normalized features are obtained by an appropriate shift in the universal-warped domain. If we use log warping given in Eq. (4), then this corresponds to a linear-scaling model of Eq. (1). The only difference between the conventional VTLN and this method is that we use WOSA for spectral smoothing and that normalization is done by appropriate shifts rather than by using the frequency-scaling method of Kamm *et al.* (1995) and Lee and Rose (1998). On the other hand, using the universal-warping function of Eqs. (20a) and (20b) would correspond to using the affine model. The only difference between Eqs. (20a) and (20b) is in the small difference in κ values obtained using the two different databases. We also see that the universal-warping functions given in Eqs. (23a) and (23b) are very similar to the affine universal-warp functions, except that they have been empirically obtained using the model of Eq. (22). In all these cases, the universal-warped spectrum is appropriately shifted for normalization, then log compressed (in magnitude), and finally a DCT is applied to derive normalized features.

C. Base line speech recognizer

The digit recognizer is developed using the hidden Markov model toolkit. The recognition of digit string is considered as a task without restricting the string length. Eleven digit models are generated for 1–9, *zero*, and *oh*. The digits are modeled as whole word hidden Markov model (HMMs) with the following parameters: 16 states per word, simple left-to-right models without skips over states, and a mixture of five Gaussian with diagonal covariance matrices per state. The silence is modeled separately using two models. The first one called *sil* consists of three states with a transition structure as suggested in Hirsch and Pearce (2000). This HMM models the silences before and after the utterance and uses a mixture of six Gaussian models per state. The second silence model called *sp* is used to model pauses between words. It consists of a single state, which is tied with the middle state of the first silence model. A 39-dimensional feature vector comprising normalized log energy, C_1-C_{12} (excluding C_0) base cepstral coefficients, and their first and second-order derivatives is used. Finally cepstral features are liftered and cepstral mean subtraction is performed, as followed conventionally.

D. Estimation of normalization factor

In ASR, there is no concept of a reference or "golden" speaker. Therefore, conventionally the estimation of normalization factor is done using a maximum likelihood (ML) framework by comparing it with the statistical model built using data from all speakers. In conventional VTLN, the linear-scaling model of Eq. (1) is used, and the scale factor α_{rs} for every speaker is estimated using the maximum likelihood criterion. In the case of shift-based normalization, the use of universal-warping function results in the spectral envelopes being translated versions of one and another in the warped domain. Speaker normalization is done by an appropriate shift in the warped domain (Sinha and Umesh, 2002, 2008). We estimate this shift factor using a maximum likelihood framework using a method analogous to the ML estimation of scale factor in conventional VTLN. We briefly describe the procedure.

Let $X_i^{\varsigma}=\{\mathbf{x}_{i,1}^{\varsigma},\ldots,\mathbf{x}_{i,T}^{\varsigma}\}$ be the set of T feature vectors obtained after the ς-shift in the universal-warped domain for an utterance from speaker i. Note that each vector $\mathbf{x}_{i,t}^{\varsigma}$ is the feature vector obtained for the tth frame of speech after applying a shift of ς_i in the universal-warped domain. Let W_i denote the transcription of the utterance from speaker i. The transcription W_i is necessary for aligning the models with the acoustic data. During testing, this transcription is obtained by doing first-pass recognition using un-normalized features and the statistical model. If ϕ denotes a set of given HMM models trained from a large population of speakers, then the optimal translation factor $\hat{\varsigma}_i$ for speaker i is obtained by maximizing the likelihood of the shifted utterances with respect to the model and the transcription, i.e.,

$$\hat{\varsigma}_i = \arg\max_{\varsigma} \Pr(X_i^{\varsigma}|\phi, W_i). \tag{24}$$

The optimal translation factor is obtained by doing a grid search over a range of values. In our experiments, we search over a range of seven shifts between −3 and +3. The method in Lee and Rose (1998) involves a grid search of scale factor α over $0.88 \leqslant \alpha \leqslant 1.12$ in steps of 0.04.

E. Recognition performance

We now discuss the recognition performance using our proposed affine model and the corresponding universal warping function given in Eqs. (20a) and (20b). We will compare it with the linear model whose universal-warping function is the log-warping function given in $\lambda=\log(f)$. We will also compare its performance to empirically obtained universal-warping function (discussed in Sec. III E) given by Eqs. (23a) and (23b). For all these models, we use shift-based normalization, and the performances are compared in a common framework. We also show the performance of the conventional MFCC, which uses linear-model and filter-bank smoothing.

We now discuss the recognition setup. When there is *no* normalization the features are computed *without* shifting in the universal-warped domain. For each such universal-warping function, we first build the HMM using un-normalized cepstral features of the training speakers. This is

TABLE II. Word error rate of various frequency-warping functions on a digit recognition task. E_b and E_n represent the word error rates before normalization (base line) and after normalization, respectively. The number in brackets refers to the equation number of the universal-warping function used.

Warping function	Scaling	Adults		Children	
		E_b	E_n	E_b	E_n
Conventional MFCC	Linear	3.48	2.98	14.38	9.35
λ [Eq. (4)]	Linear	3.24	2.85	13.04	9.20
ν_{PnB} [Eq. (20a)]	Nonlinear	2.97	2.56	13.49	8.00
ν_{HiL} [Eq. (20b)]	Nonlinear	3.04	2.57	13.42	8.02
τ_{PnB} [Eq. (23a)]	Nonlinear	3.09	2.53	13.59	7.88
τ_{HiL} [Eq. (23b)]	Nonlinear	3.03	2.49	13.60	7.77
$\eta_{\kappa=700}$ [Eq. (12)]	Nonlinear	3.02	2.52	13.73	7.96

referred to as the *base line* system. To evaluate the performance during normalization, it is necessary to train the HMM with the normalized features from the training speakers. For each universal-warping function, we estimate the ML shift factor for the training speakers with respect to the base line model for that universal-warping function using Eq. (24). We then do the appropriate shift in the universal-warped domain and compute the corresponding normalized features. The *normalized* HMM model is built from the normalized features of the training speakers. This is also the procedure followed in all conventional VTLN implementation methods since there is no "reference" or golden speaker for whose speech the HMM model can be built. Using the first-iteration normalized HMM model, we again estimate the ML shift factor for the training speakers. Using the new estimates of shift, we again compute normalized features, which are then used to build the second-iteration normalized HMM. This process is iterated three times to refine the normalized HMM model (Lee and Rose, 1998). During testing, the shift factor for each test speaker is again estimated using the maximum likelihood criterion. Similar to conventional VTLN, the transcription, W_i, for speaker-factor estimation during testing is obtained from the first recognition pass using un-normalized features and the base line model. After appropriate normalization using the estimated shift factor, the corresponding features of the test speaker are used for recognition.

Table II shows the word error rates before normalization (base line), E_b, and after normalization, E_n. The first row marked "Conventional MFCC" is obtained by using the conventional filter-bank front-end and the normalization is done using the linear-scaling model [see Lee and Rose (1998)]. All the other results are obtained under the common framework of shift-based normalization with WOSA spectral smoothing. In the shift-based normalization scheme, we need to use an appropriate universal-warping function for each model. We compare the performance of the following models:

(i) Linear-scaling model whose corresponding universal-warping function is the log-warping function $\lambda = \log(f)$.
(ii) The proposed affine model in Eq. (7) whose universal-warping function ν_{PnB} and ν_{HiL} is given by Eqs. (20a) and (20b) for PnB and HiL, respectively. Note that these two universal-warp functions are very similar with κ in the neighborhood of 500. We have shown recognition results for both to illustrate the fact that small differences in κ values do not affect the performance significantly.
(iii) The empirically determined universal-warp functions, τ_{PnB} and τ_{HiL} of Eqs. (23a) and (23b) obtained by fitting the piecewise approximation to the universal-warp function of Umesh *et al.* (2002b, 2002c).
(iv) Finally, note that all the nonuniform scaling models have universal-warping functions that are parametrically similar to the mel scale. In speech recognition the mel scale is given by Eq. (12), with b (or κ) being 700. Therefore, we have also compared the performance using this universal-warping function, $\eta_{\kappa=700}$, with $\kappa=700$. Note that this universal-warp function ($\eta_{\kappa=700}$) corresponds to the affine model of Eq. (7) with $\kappa=700$.

Some important observations from Table II are as follows:

(1) The shift-based normalization scheme along with WOSA front-end performs significantly better than the conventional MFCC front-end with scale-based normalization. This can be seen from the top two rows where we compare the performance of λ with conventional MFCC both of which correspond to uniform/linear-scaling model. In Umesh *et al.* (2004) and Sinha and Umesh (2008), we discuss the reasons for this improvement.
(2) All nonlinear normalization models perform significantly better than conventional linear-scaling model. This can be seen by comparing the performance of the $\lambda=\log(f)$ warp with the nonlinear warp functions. Note that all these methods have used WOSA front-end.
(3) All the nonlinear models have a similar functional form for the universal-warp function, with only κ varying. Since they all have similar performances, it may be concluded that small changes in κ value do not significantly affect recognition performance.

We summarize from Table II that nonuniform normaliza-

tion is more appropriate for speaker normalization than for uniform normalization. More importantly, we have a parametric affine model to describe the nonuniform relationship between speakers.

This study also shows the normalization performance for different values of b for the warping function in Eq. (11). The value of $b=641.94$ seems to be marginally better than $b=700$ in terms of recognition performance, which is interestingly closer to $b=657.6$ as computed in Umesh *et al.* (1999) for the Stevens and Volkman (1940) data.

V. DISCUSSION AND CONCLUSION

We have proposed a parametric model that affinely relates the formant frequencies of any two speakers enunciating the same sound and have shown the corresponding universal-warping function to be similar to the mel scale. We also justified the validity of this model by showing a detailed comparison to other models in terms of formant fitting error and recognition performance. Hence, we claim that the proposed affine model is a more appropriate model for nonuniform speaker normalization than the conventional uniform scaling and other *ad hoc* nonuniform models.

The other important aspect of the paper is in showing a possible connection between speaker normalization and psychoacoustics. We claim so because the proposed warping function—computed from speech data alone—behaves very similar to the mel scale, which has been obtained from psychoacoustic studies. While the mel scale is used widely in speech recognition, we believe that this is the first time that it has been shown that the mel scale is also important for speaker normalization.

ACKNOWLEDGMENTS

We would like to thank Professor Terrance Nearey for reading the manuscript carefully and giving us many useful comments and suggestions, especially with respect to the material in Sec. III. We would also like to thank the anonymous reviewers for their suggestions, which have helped to improve the manuscript. This work was supported in part by the Department of Science and Technology, Ministry of Science and Technology, India under SERC Project No. SR/S3/EECE/0008/2006.

Acero, A., and Stern, R. M. (**1991**). "Robust speech recognition by normalization of the acoustic space," in Proceedings of IEEE ICASSP, Toronto, Canada, pp. 893–896.

Bladon, R. A. W., Henton, C. G., and Pickering, J. B. (**1983**). "Towards an auditory theory of speaker normalization," Language and Communication0271–5309 **4**, 59–69.

Eide, E., and Gish, H. (**1996**). "A parametric approach to vocal tract length normalization," in Proceedings of IEEE ICASSP '96, Atlanta, USA, pp. 346–348.

Fant, G. (**1975**). "A non-uniform vowel normalization," Technical report, Speech Transmission Laboratory, Royal Institutte of Technology, Stockholm, Sweden.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K. (**1995**). "Acoustic characteristics of american English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hirsch, H. G., and Pearce, D. (**2000**). "Aurora framework for the performance evaluation of speech recognition systems under noisy condition," in ISCA ITRW ASRU 2000, Automatic Speech Recognition: Challenges for the New Millenium.

Kamm, T., Andreou, G., and Cohen, J. (**1995**). "Vocal tract normalization in speech recognition: Compensating for systematic speaker variability," in Proceedings of the 15th Annual Speech Research Symposium, Johns Hopkins University, Baltimore, pp. 175–178.

Lee, L., and Rose, R. C. (**1998**). "A frequency warping approach to speaker normalization," IEEE Trans. Speech Audio Process. **6**, 49–59.

McDonough, J., Bryne, W., and Luo, X. (**1998**). "Speaker normalization with all-pass transforms," in Proceedings of ICSLP '98, Sydney, Australia.

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2134.

Nearey, T. M. (**1978**). "Phonetic feature systems for vowels," Technical report, Indiana University Linguistics Club.

Nearey, T. M. (**1992**). "Applications of generalized linear modeling to vowel data," in Proceedings of ICSLP '92 Canada.

Nordström, P. E., and Lindblom, B. (**1975**). "A normalization procedure for vowel formant data," in International Congress on Phonetic Science, Leeds, England.

Nuttall, A. H., and Carter, G. C. (**1982**). "Spectral estimation using combined time and lag weighting," Proc. IEEE **70**, 1115–1125.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Potamianos, A., and Narayanan, S. (**2003**). "Robust recognition of children's speech," IEEE Trans. Speech Audio Process. **11**, 603–616.

Sinha, R., and Umesh, S. (**2002**). "Non-uniform scaling based speaker normalization," in ICASSP '02, Orlando, pp. 589–592.

Sinha, R., and Umesh, S. (**2008**). "A shift-based approach to speaker normalization using non-linear frequency-scaling model," Speech Commun. **50**, 191–202.

Stevens, S. S., and Volkman, J. (**1940**). "The relation of pitch to frequency: A revised scale," Am. J. Psychol. **53**, 329–353.

Syrdal, A. K., and Gopal, H. S. (**1983**). "Perceived critical distances between F_1-F_0, F_2-F_1, F_3-F_2," J. Acoust. Soc. Am. **74**, S88–S89.

Umesh, S., Cohen, L., and Nelson, D. (**1999**). "Fitting the mel scale," in Proceedings of IEEE ICASSP '99, pp. 217–220.

Umesh, S., Cohen, L., and Nelson, D. (**2002a**). "Frequency warping and the mel scale," IEEE Signal Process. Lett. **9**, 104–107.

Umesh, S., Cohen, L., and Nelson, D. (**2002b**). "The speech scale," ARLO **3**, 83–88.

Umesh, S., Kumar, S. V. B., Vinay, M. K., Sharma, R., and Sinha, R. (**2002c**). "A simple approach to non-uniform vowel normalization," in Proceedings of IEEE ICASSP '02, Orlando, USA, pp. 517–520.

Umesh, S., Sinha, R., and Kumar, S. V. B. (**2004**). "An investigation into front-end-signal processing for speaker normalization," in Proceedings of IEEE ICASSP'04, Montreal, USA.

Wakita, H. (**1977**). "Normalization of vowels by vocal-tract length and its application to vowel identification," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-25**, 183–192.

Wegmann, S., McAllaster, D., Orloff, J., and Peskin, B. (**1996**). "Speaker normalization on conversational telephone speech," in IEEE ICASSP '96, Atlanta, USA, pp. 339–341.

Zhan, P., and Waibel, A. (**1997**). "Vocal tract length normalization for large vocabulary continuous speech recognition," Technical report, School of Computer Science, CMU, Pittsburgh, USA.

Zhan, P., and Westphal, M. (**1997**). "Speaker normalization based on frequency warping," in Proceedings of IEEE ICASSP '97, Munich, Germany, pp. 1039–1042.

Modeling the temporal dynamics of distinctive feature landmark detectors for speech recognition

Aren Jansen and Partha Niyogi[a)]
Department of Computer Science, University of Chicago, 1100 East 58th Street, Chicago, Illinois 60637

(Received 13 September 2007; revised 10 June 2008; accepted 11 June 2008)

This paper elaborates on a computational model for speech recognition that is inspired by several interrelated strands of research in phonology, acoustic phonetics, speech perception, and neuroscience. The goals are twofold: (i) to explore frameworks for recognition that may provide a viable alternative to the current hidden Markov model (HMM) based speech recognition systems and (ii) to provide a computational platform that will facilitate engaging, quantifying, and testing various theories in the scientific traditions in phonetics, psychology, and neuroscience. This motivation leads to an approach that constructs a hierarchically structured point process representation based on distinctive feature landmark detectors and probabilistically integrates the firing patterns of these detectors to decode a phonological sequence. The accuracy of a broad class recognizer based on this framework is competitive with equivalent HMM-based systems. Various avenues for future development of the presented methodology are outlined.
 [DOI: 10.1121/1.2956472]

PACS number(s): 43.72.Ne [DOS] Pages: 1739–1758

I. INTRODUCTION

We are interested in the problem of *pure speech recognition*—the ability of humans to interpret (decode) the speech wave form reaching their ears in terms of a sequence of phonological units without invoking any higher level (syntactic, semantic, or pragmatic) linguistic knowledge. This ability to perform pure acoustic–phonetic decoding is manifest in our ability to recognize streams of nonsense words, to detect familiar words in unfamiliar streams,[1] and in a host of other such basic phonological perception processes. Furthermore, this ability must underpin all higher level language learning; after all, every time we hear a new word, it is effectively a nonsense word for us. We are able to perform such acoustic–phonetic decoding in the absence of higher level linguistic knowledge, but we do need knowledge of phonetic, phonotactic, and phonological regularities of the language at hand. Correspondingly, our method rests heavily on accurate models of such regularities.

Our approach is based on the following principles:

(1) In our system, *distinctive features* are the atomic units in terms of which all higher level phonological units such as broad phonological classes, phonemes, syllables, and the like are composed. This is in contrast to most traditional speech recognition systems that use phone-based units (e.g., triphones). The distinctive features are symbolic (discrete), define natural classes, have articulatory and acoustic correlates, are hierarchically structured, and may overlap in time as suggested by autosegmental phonology.[2–4] They also provide a compact representation for phonological rules to capture coarticulatory effects that result in pronunciation variability.
(2) Distinctive features are associated with articulatory gestures that have natural acoustic consequences in the speech signal. The acoustic properties that distinguish different sounds from each other and different distinctive features from each other reside at different scales in time and frequency. Consequently, rather than having a "one size fits all" representation as is common in much of modern speech recognition, we will have multiple representations tuned for different distinctions.
(3) Correspondingly, we build *feature detectors* for the patterns and characteristic signatures for the distinctive features. Of particular significance in our system is the points in time when feature detectors fire. These points in time are associated with important events or *landmarks*, which correspond to maxima, minima, and inflection points of specialized acoustic properties. As a result, one obtains a *sparse, point process* representation of the speech signal. Such a representation is reminiscent of spike train patterns observed in the behavior of selective neurons, particularly in parts of the auditory cortex of a variety of animals (see Refs. 5 and 6). Such a spike-based representation is in contrast to the vector time series (i.e., frame-based) representation used in nearly all modern recognition systems.
(4) The decoding of the signal proceeds by integrating the firings of the individual feature detectors in a hierarchical way, where high-level decisions trigger off further context-dependent processing at lower levels. At the root of the hierarchy is the sonorant-obstruent feature, which is the most basic and perceptually salient distinction among speech sounds. Vowels correspond to peaks of the sonority profile and provide anchor points that define syllable-sized analysis units. Integration of detector outputs occurs at such *syllabic* time scales on the rationale that this is the smallest perceptually robust unit. The information content of the signal within each analysis unit is coded in the temporal statistics of the point process

[a)]Electronic mail: niyogi@cs.uchicago.edu

representation. This is in the spirit of *temporal coding* in neural systems and allows us to model durational patterns in a novel way.

Our goal in this paper is to give computational expression to these principles. In recognition of the inherent variability in the acoustic correlates of distinctive features, we pay particular attention to what might be natural and coherent statistical frameworks in which to model different aspects of the signal. In the process, we end up with a system that is quite unlike any other built so far, though it is clearly closest in spirit to those inspired by acoustic phonetics, distinctive features, and landmarks (e.g., Refs. 7–10; see Sec. II E for discussion).

One may ask, what are the benefits of an approach such as ours? First, our motivations may be traced more directly to scientific understanding of related phenomena in linguistics, psychology, and neuroscience. Consequently, our system may provide a computational platform to test specific hypotheses about speech perception in these fields. Second, the simplicity of our modular design may aid diagnostics and portability to new languages and environments. Third, the hierarchical approach leads to fewer parameters, allows reuse of training data for different distinctions, and makes possible the efficient training of our system from limited amounts of training data. Finally, the system design, with its specialized detectors and temporal coding, provides a new way to characterize the statistics of speech signals and reason about issues of invariance and robustness.

As an intermediate step toward our long-term goal, we have chosen in this paper to concentrate on the task of broad class recognition. This is the simplest nontrivial sequence recognition task, requiring not only the capacity to distinguish between phonological classes (i.e., classification), but also manage possible insertion and deletion errors. For this reason, broad class recognition has often been an intermediate step in evaluating new approaches (see Refs. 8 and 11). It is worth noting that broad class recognition is not without practical merit in its own right. As one example, Huttenlocher and Zue (see Ref. 12) proposed an approach to lexical access based on partial phonetic information (i.e., broad class information only). A method of incorporating a broad class recognizer in noisy environments to improve robustness has also been proposed (see Ref. 13). Finally, certain small vocabulary tasks can be performed with a broad class recognizer. One straightforward example is spoken digit recognition, as each digit has a unique broad class sequence (see Ref. 14). Another possibility would be key-word spotting, where partial phonetic information could be used for a low specificity search.

In Sec. II, we outline the overall architecture of the system. In Secs. III and IV, we elaborate on the computational details and experimental performance of each of the modules. It is worth noting that the approach of this paper is only a particular instantiation of the general principles outlined earlier and we discuss variants at each stage. Consequently, in Sec. V we discuss future directions and in particular the challenges to work up to a complete recognition system.

II. OVERALL SYSTEM ARCHITECTURE

Here we elaborate on the principles outlined in Sec. I, motivate them, describe how they are instantiated in our overall architecture, and make connections to related research at appropriate points.

A. Distinctive feature representation

Our goal is to build a machine that takes as input the speech signal and produces as an output a symbolic (linguistic) representation. It is overwhelmingly the case that modern speech recognition systems based on hidden Markov models (HMMs) use some kind of phone-based representation (typically triphones in large scale applications) with an HMM mediating the mapping between phone sequences and an acoustic vector time series.

In contrast, we will represent the phonological units in terms of distinctive features. Since this may be a confusing term to the speech engineer (in the terminology of the engineer or statistician, a feature is typically a property of the speech signal), it is worthwhile to elaborate a little bit on the phonological notion of distinctive feature, its justifications, and its implications for our approach.

Although antecedents of the feature may be found in the sivasutras of Panini, the modern notion is usually traced to Trubetzkoy, Jakobson, and others in the early part of the 20th century. While we recognize intuitively that the objects of language may be hierarchically composed of smaller objects (at a rough cut, sentences are made up of phrases, phrases of words, words of syllables, and syllables of phonemes), it was the insight of Jakobson and others that phonemes were not the primitive, atomic units or building blocks of language but rather each phoneme may be usefully considered to be a complex of distinctive features. These features typically emphasized binary oppositions among groups of phonemes that share some common characteristic. Thus a collection $f_1, \ldots, f_k$ of binary-valued features define a possible set of at most 2^k different phonemes in a language.[15]

Distinctive features are motivated by several different considerations. First, distinctive features provide one way to group the phonemes into *natural classes*. The primary argument comes from linguistics, where it is a curious fact of language that some groups of phonemes seem to behave as an equivalence class as they participate in phonological processes. For example, in English, to pluralize a noun, one takes the root word (stem) and modifies it as follows:

If the word ends in /s z tʃ dʒ ʃ ʒ/ add /ɪz/,
  Examples: places, porches, cabbages, ambushes;
else if the word ends in /p t k f θ/ add /s/,
  Examples: lips, lists, maniacs, telegraphs;
else add /z/,
  Examples: clubs, herds, fangs, holes, pies
(after Ref. 16).

Thus, the sounds /p t k f θ/ behave as a class with respect to this rule. When one examines the class of sounds /p t k f θ/ and asks what they share in common that distinguishes them from the subclass /b d g v ð/, one arrives at the understanding that the former consists of unvoiced sounds while the latter consists of their voiced counterparts. Indeed,

TABLE X. Confusion matrix for $S/N=+12$ db and frequency response of 200–1200 cps.

	p	t	k	f	θ	s	ʃ	b	d	g	v	ð	z	ʒ	m	n
p	165	46	31	3	1				1				1			
t	91	83	68	4	1	2			1			2				
k	48	55	147	2	3							1				
f	16	4	3	146	60	3	2	11			1	2				
θ	4	3		109	76	17	2	12	1			2	1	1		
s	2	1	1	43	83	83	11	3		1	1	7				
ʃ	1	6	2	12	41	86	90		6	4		4				
b				14	5			223	4		5	1				
d	1				1	3	4	4	173	37		2	1	2		
g	1					1			102	107	1	2	7	7		
v	2	2		2	1			23	1	2	163	62	14	3	1	
ð				1		3	2	27	6	32	87	107	36	7		
z	1							4	12	48	10	15	114	39		1
ʒ							1		3	35	1	16	60	134	2	
m	1									1					229	9
n															5	247

FIG. 1. Consonant confusion matrix from Ref. 19.

voicing is the feature distinguishing the minimal pairs (/b p/, /t d/, /k g/, /f v/, /θ ð/) with phonemes /p t k f θ/ being [−voiced] and phonemes /b d g v ð/ being [+voiced]. Indeed, the English plural rule may be summarized by the following: If the word ends in [+coronal +strident], add /ɪz/; if it ends in [−voiced +stiff vocal cords] add /s/; add /z/ otherwise. A large number of similar examples in English may be found in Ref. 17, while subsequent works have elaborated and refined the analysis for many different languages of the world (see Ref. 18 for some details).

A second argument comes from studies of phonetic perception in psychology. For example, the classic work of Miller and Nicely[19] tested various subjects on their ability to discriminate between consonantal phonemes at different noise levels. Confusion matrices were constructed and shown in Fig. 1 is an example of one such confusion matrix at 12 dB signal-to-noise ratio. Notice the block diagonal structure of the confusion matrix with a natural grouping of phonemes into perceptually similar classes. These coincide with natural classes organized by distinctive features.

A third argument comes from articulatory and acoustic correlates of the phoneme classes. Following principles of speech production, phonemes produced with similar articulatory gestures will have similar acoustic properties. Since all linguistically relevant sounds (phonemes) are produced by the manipulation of a small number of articulators (the tongue, lips, glottis, nasal coupling, etc.), one may group phonemes into classes based on similarity of articulatory configurations. Thus, the distinction between the groups of vowels /i ɪ u/ and /a ɑ æ/ can be based on feature of tongue height (high versus low); the distinction between /p b m/ and /t d n/ can be based on place of constriction (labial versus coronal) in the vocal tract when producing the consonants. Reference 20 is an elaboration of this aspect of speech analysis.

The last few decades have seen a convergence of these kinds of arguments into different kinds of feature systems that differ from each other in detail based on their application purpose but have the same natural coherence. Thus, feature systems are rooted in phonology but have natural articulatory interpretations and corresponding acoustic and perceptual correlates.

More recently, work by Goldsmith,[2] Sagey,[3] McCarthy,[4] and others suggests that features have an internal organization with a hierarchical relationship with respect to each other. For our purposes, we will adopt the hierarchy of Fig. 2 leading to the major classes of vowels, approximants, nasals, stops, fricatives, and affricates. This hierarchy involves the following distinctive features:

(1) [son]: distinguishes sonorant sounds ([+son]) that are made with an open vocal tract (such as vowels, approximants, and nasals) from obstruent sounds ([−son]) such as stops and fricatives.
(2) [cons]: distinguishes consonantal sounds from vowel sounds.
(3) [cont]: distinguishes stop consonants [−cont] from everything else. [−cont] sounds are produced with a complete closure of the airway at some point during the articulation.
(4) [nasal]: distinguishes those sounds that couple the nasal cavity /n m ŋ/ from the rest.

These features are closely related to manner features,[21] stricture features,[18] and articulator-free features.[20]

What is the significance of the hierarchy in our context? The hierarchy has a justification in our minds from several perspectives.

(1) Nodes higher up in the tree correspond to features that are somehow more basic or fundamental and whose acoustic correlates are less context dependent. As a result, these features may be derived from the speech signal in a more robust and reliable way.
(2) Some features are irrelevant based on values of certain other features. For example, if a sound is [−son], then

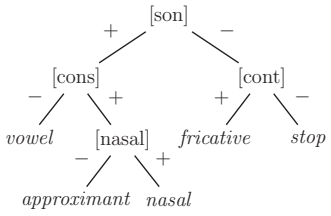


FIG. 2. The hierarchy of distinctive features leading to the broad (manner) classes.

the sound is automatically [−nasal]. Similarly, if the sound is [+son], then the sound is automatically [+cont]. Further features come into play once the broad class distinctions are made.

(3) The grouping of the features indicate a further restriction in the set of natural classes that are relevant for phonological processes.

As a result of the above-presented observations, we see that a featural notational system allows a more compact description of phonological processes such as assimilation and coarticulatory dynamics. Thus, pronunciation networks expressed featurally are likely to be more streamlined than those that are expressed in phonemic units.

So far, we have discussed how distinctive features allow one to describe the set of phonemes without regard to the fact that linguistic utterances are composed of words that are realized as linear sequences of segmental units or phonemes. A traditional view within generative phonology regards each phoneme to be a bundle of distinctive features and therefore each sequence is a sequence of feature vectors (bundles). This suggests an idealized view of a phonological utterance as a sequence of segments with the distinctive feature values being synchronized in time. However, we will adopt a representation that is more in line with the developments of autosegmental phonology, where associated with each node of the hierarchy we have a timing tier showing how the particular feature is distributed in time. Since the distinctive feature is a binary-valued, phonological variable, it will need to be estimated from the acoustic properties of the speech signal. Associated with these features are natural articulatory gestures with their corresponding acoustic consequences. These acoustic consequences will define points in time that are naturally related to the realization of the feature in time.

B. The sonority profile and vowels

The most obvious and perceptually salient aspect of the speech signal is the sonority profile of the signal distinguishing the sonorant sounds from the obstruent; recent psychological studies support this claim.[22] At the same time, perhaps the most basic distinction between classes of phonemes in their role in phonological processes is that between the consonants and the vowels. These two are related through the notion of a sonority hierarchy[23] with vowels occupying the top of the sonority hierarchy as the most open, full throated sonorant class of sounds. As a result, vowels correspond to local maxima of the sonority profile, occupy the nucleus position of syllables, and are the stress-bearing elements in the intonational contours of the speech stream.

Consequently, our entry point into the interpretation of the signal for further analysis is to

(1) Segment[24] the signal into sonorant and obstruent regions.
(2) Identify the vowel landmarks as the peaks of the sonority profile.

It is worthwhile to make a few remarks about the sonorant–obstruent distinction and its differences from the consonant–vowel distinction. Intuitively, sonorant sounds are produced with a relatively open vocal tract, with no significant pressure buildup anywhere in the tract, and with the vocal cords vibrating. Consequently, such sounds are always voiced; more significantly, from our point of view, because these sounds are associated with vibrating vocal cords, they are always periodic. The lack of significant pressure buildup and the relatively open vocal tract results in these sounds having marked formant structure in the lower frequency bands and relatively high energy. Thus sonorant sounds are easy to distinguish acoustically from obstruent sounds: This is perhaps the easiest distinction to make and as a result, we place this feature at the top of our hierarchy.

While all obstruent sounds are consonantal, not all sonorant sounds are vocalic. The approximants and nasals are consonantal sounds that are sonorant. Acoustically they are very similar to vowels, yet, phonologically, they play the role of consonants. The distinction between consonants and vowels is tied intimately with the syllable structure of a phonological sequence. The number of syllables in an utterance is equal to the number of vowels in that utterance with each vowel occupying the nucleus of a syllable.[25] Following the sonority hierarchy, and noting that vowels are at the top of the hierarchy, one finds that syllabic nuclei coincide with the local maxima of the sonority profile of the utterance.

Since we begin by identifying the sonorant and obstruent regions and the vowel landmarks,[26] the task of modeling the utterance reduces to modeling the intervocalic segments. The intervocalic segments correspond to continuous sequences of consonants that lie between two vowels. These are made up of sequences of sonorant consonants and obstruent consonants that are modeled separately. By doing this, we model portions of the signal at a syllabic scale. This potentially allows us to capture coarticulatory effects. Following the arguments of Ref. 27, there seems to be some justification in this choice from perceptual considerations as well.

C. Feature detectors for subsequent processing

The distinctive features corresponding to the lower tiers of Fig. 2 make further distinctions. Associated with each such feature, we build a *feature detector* that ideally should detect the presence or absence of that feature. One has to come to terms with the basic fact of speech that while the distinctive feature is categorical, the acoustic correlates are continuous and gradient. Consequently, the output of the feature detector is a real-valued function of time whose magnitude at each time may sometimes be interpreted as the degree (probability) to which the distinctive feature is present at that time.

There are three key ideas involved in the construction of a feature detector. First, for each feature of interest, a specialized acoustic representation is constructed in which that feature best expresses itself, i.e., a representation that captures the acoustic correlates of the feature that help distinguish positive from negative instantiations. Second, in this representational space, a classifier is built to separate positive examples from negative ones.[28] Third, the real-valued output of the classifier is further processed to provide a sparse rep-

resentation in time as a *point process*. Thus, maxima, minima, and inflection points of the output of the classifier correspond to points in time when "interesting" phonological events take place. Following the general philosophy of Stevens[7] and others, we refer to these points in time as *landmarks*.

Let us describe briefly four detectors that may be associated with leaf nodes of the distinctive feature tree of Fig. 2. These effectively serve as broad (manner) class detectors at each point in time.

(1) **Stop detector:** This is a detector for the plosive sounds /p t k b d g/. These sounds are [−cont] and are produced when there is a complete closure of the vocal tract followed by a sudden release. This leads to the acoustic consequence of having a period of extremely low energy followed by a noisy, wideband burst spectrum. Thus one is looking for a transition from closure to burst in the signal. Details surrounding the construction of a stop detector are provided in Refs. 29 and 30. We implement a variant of such a detector in our system.

(2) **Fricative detector:** This is a detector for the class of fricated sounds /s z ʃ ʒ f v θ ð/. These sounds are produced with a partial closure of the vocal tract so that there is turbulent pressure generated, leading to a noisy signal whose broad spectral profile is governed by the shape of the vocal tract. One may try to make a detector that fires during the fricated regions of the signal, defining landmark locations that correspond to local maxima of the continuous classifier output.

(3) **Nasal detector:** This is a detector for the nasal sounds /n m ŋ/ that are produced with a coupling of the nasal cavity. This nasal coupling has some characteristic acoustic consequences: The total energy is reduced substantially from the neighboring vowel, there is a drop-off in energy above 500 Hz, and the first formant is around 300 Hz. Here, a detector may be constructed to locate landmarks when nasal coupling is maximal.

(4) **Approximant detector:** This detector for liquids /l r/ and glides /h j w/ is the most difficult to implement because their acoustic characteristics are not completely distinct from adjacent vowels. Formant transitions and energy profiles in appropriately chosen bands may provide some discriminatory power.

Construction of each of the above-presented detectors constitute research projects in their own right (see Refs. 31–33 for examples). Our stop detector is constructed with an acoustic representation motivated directly by speech production. We construct the other detectors by computing canonical short-time (windowed) mel frequency cepstral coefficients (MFCCs) with window lengths, frame rates, and frequency ranges that are appropriate for each of the broad classes in question. Specifics regarding detector construction are provided in Secs. III B and IV B. Some further remarks are worthwhile:

(1) Sonorant regions are fundamentally different from obstruent regions of the signal. In sonorant regions, the signal is periodic and formant structure is evident. Most of the information content is at low frequencies and measures of spectral gravity, formants, and the like make sense. Consequently, the construction of nasal and approximant detectors and all further context-dependent processing may need to be sensitive to such considerations. We leave such refinements to future work.

(2) We do not pay attention to the output of the feature detector at all points in time. This leads to a sparse representation that may have some computational advantages. The specific points in time may be interpreted as the most relevant from perceptual or articulatory perspectives. These points may be related to the notion of landmarks in the theory of Stevens.[7] The duration between these points in time is naturally correlated with the timing of articulatory and perceptual events and may be modeled directly in our framework. Thus, rather than model the detailed value of the feature detector output at all points in time, we shift the burden to the modeling of the durational statistics.

(3) Our detectors may also be viewed as complex feature detectors that may themselves be trained on the output of more basic detectors along the lines of Ref. 34. It is also worthwhile to reflect on various neurophysiological findings that suggest the presence of neurons that fire selectively when certain complex acoustic attributes are present in the input stimulus. Our detectors may be analogized to such selective neurons.[5] A further connection to neurobiologically motivated speech recognition models may be found in Ref. 35, where spike pattern recognition models were also suggested.

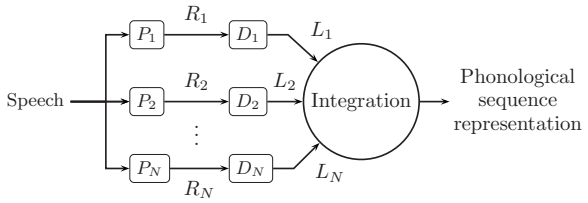


FIG. 3. Architecture of our landmark-based system. Here, $\{P_i\}$ are signal processors that output representation $\{R_i\}$, which are input into a set of feature detectors, $\{D_i\}$. The detectors output a set of candidate landmarks, $\{L_i\}$, which are probabilistically integrated to predict a phonological sequence.

At this point, the architecture of the system looks like that shown in Fig. 3. The speech signal is processed by several different signal processing transformations $(P_1,\ldots,P_N)$ to give rise to multiple representations $(R_1,\ldots,R_N)$. In each representational space, a feature classifier acts producing a real-valued output (first stage of $D_1,\ldots,D_N$). The last step is picking maxima, minima, or inflection points in the output of the classifier (second stage of $D_1,\ldots,D_N$) giving rise to points in time where the feature is most acoustically or perceptually prominent. These points in time may be naturally associated with events or landmarks $(L_1,\ldots,L_N)$, around which further context-dependent processing may be conducted.

Because the distinctive features have an internal hierarchical structure, we end up with a representation of speech as a collection of marked point processes associated with the

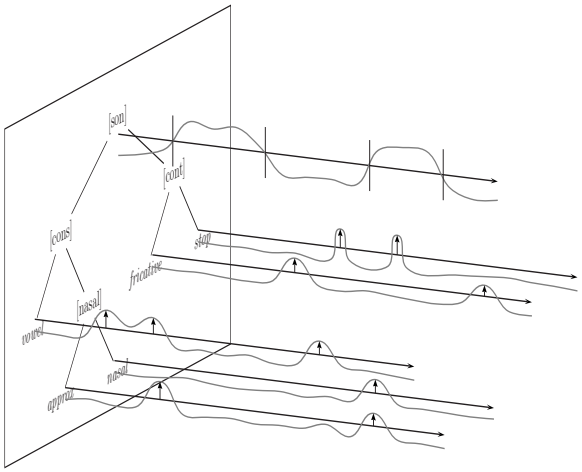


FIG. 4. Schematic diagram of the hierarchical timing tier representation. Landmarks are indicated by vertical arrows.

hierarchically structured timing tiers as shown in Fig. 4. A few aspects of this representation are worth noting. First, we have access not just to the points in time when the detector output peaks but also to the strengths of the detectors at that point, resulting in a natural marked point process. Second, the representation is very sparse. This is in stark contrast to typical representations of speech where one models the local spectral content at every time instant leading to a vector time series. Third, the information in the representation is now coded in the temporal dynamics of these spikes and it is natural to expect that the statistics of interspike times will be correlated with the durations between articulatory events and ultimately the durations of various linguistic segments. Finally, because of potentially different processing at different tiers, the time scales at which the different feature detectors fire may be quite different. This representation thus reflects our intuition that events in speech occur at multiple time scales that need to be decoupled from each other.

D. Integration in time

The challenge now is to integrate the firings of all these detectors in a coherent way to decode the sequence of broad classes that underlie the speech stream. In other words, we want a consistent way to map the hierarchical point process representation into a linear sequence of phonological units that are most likely to represent the underlying message in the speaker's mind.

The key idea is that we first segment the utterance into sonorant and obstruent regions and identify all the vowels in the utterance. The vowels are the nuclei of the syllables in the utterance and now each intervowel region is mapped into its most likely broad class sequence. If our feature detectors worked perfectly, this task would be trivial: Simply read off the output of the feature detectors to obtain the corresponding broad class sequence. On the other hand, we have to deal with the fact that our detectors have nonzero error rates. Thus, if between two vowel landmarks, the underlying sequence was /r k s/, we expect the stop detector fires exactly once at an appropriate point in time in between the two vowel landmarks and in an obstruent region of the signal. However, the stop detector may fire twice (for example, once correctly at the closure–burst transition and once incorrectly at some other point).

TABLE I. Definition of broad (manner) classes used in our system, in order of descending sonority.

Broad Class	Abbreviation	Sonorant	Phones (Arpabet)
Vowels	V	Yes	iy, ih, eh, ey, ae, aa, aw, ay, ah, ao, oy, ow, uh, uw, ux, er, ax, ix, axr, ax-h
Approximants	A	Yes	1, r, w, y, hh, hv, el
Nasals	N	Yes	m, n, ng, em, en, eng, nx
Fricatives	F	No	s, sh, z, zh, f, th, v, dh, jh, ch
Stops (Plosives)	P	No	b, d, g, p, t, k, dx, q
Silence	sil	No	bcl, dcl, gel, pci, tcl, kcl, pau, epi, h#

Therefore, we need to model the statistical distribution of the pattern of firings associated with each underlying sequence and choose the most likely sequence given the observed pattern of firings. Let

$$O = \{O^{\text{seg}}, O^{\text{v}}, O^{\text{sc}}, O^{\text{obs}}\}$$

be the set of observables associated with the timing tier representation, where

(1) $O^{\text{seg}} = \{0, t_1^{\text{seg}}, t_2^{\text{seg}}, \ldots, T\}$ contains the sonority segmentation boundaries,
(2) $O^{\text{v}} = \{(t_i^{\text{v}}, f_i^{\text{v}})\}$ contains the vowel landmarks,
(3) $O^{\text{sc}} = \{(t_i^{\text{sc}}, f_i^{\text{sc}})\}$ contains the sonorant consonant (approximant and nasal) landmarks, and
(4) $O^{\text{obs}} = \{(t_i^{\text{obs}}, f_i^{\text{obs}})\}$ contains the obstruent (fricative, stop, and silence) landmarks.

Here, each landmark is described by an ordered pair of the form (t, f), where t is the instance in time it occurred and f is the strength of the detection.

Given the timing tier observation variables O, our goal is to choose the most likely broad class sequence, $B \in \Sigma^*$ for $\Sigma = \{\text{V}, \text{N}, \text{A}, \text{F}, \text{P}, \text{sil}\}$ (see Table I for definitions of symbols), according to the maximum *a posteriori* (MAP) rule,

$$\begin{aligned} B_{\text{opt}} &= \arg \max_{B \in \Sigma^*} P(B|O) \\ &= \arg \max_{B \in \Sigma^*} P(B|O^{\text{seg}}, O^{\text{v}}, O^{\text{obs}}, O^{\text{sc}}). \end{aligned} \tag{1}$$

Applying Bayes' rule, and noting that $P(O^{\text{obs}}, O^{\text{sc}})$ is a constant in the optimization, we can write

$$B_{\text{opt}} = \arg \max_{B \in \Sigma^*} P(B|O^{\text{seg}}, O^{\text{v}}) P(O^{\text{obs}}, O^{\text{sc}}|B, O^{\text{seg}}, O^{\text{v}}).$$

According to the feature hierarchy, we model the terms $P(B|O^{\text{seg}}, O^{\text{v}})$ and $P(O^{\text{obs}}, O^{\text{sc}}|B, O^{\text{seg}}, O^{\text{v}})$ as follows (see Fig. 5).

(1) *Definition of sonorant and obstruent segments*: The sonority transition times O^{seg} define a set of interleaved obstruent and sonorant segments, denoted as

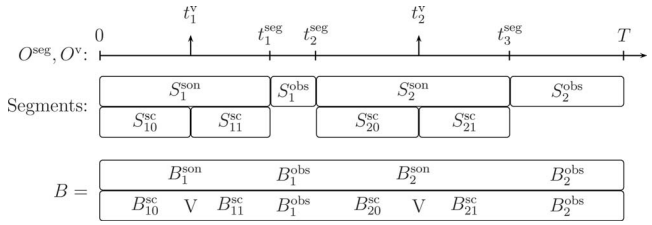


FIG. 5. Schematic representation of the segments and broad class subsequences used in the model derivation. Here we assume a sonorant segment comes first.

$$\{S_1^{\text{obs}}, \ldots, S_{N^{\text{obs}}}^{\text{obs}}\} \text{ and } \{S_1^{\text{son}}, \ldots, S_{N^{\text{son}}}^{\text{son}}\},$$

respectively, where each S_i^{obs} and S_j^{son} is a time interval. Without loss of generality, assume S_1^{son} occurs before S_1^{obs} and $N^{\text{son}}=N^{\text{obs}}$. Given a candidate sequence B, we declare it to be consistent with O^{seg} if there exist mappings $S_i^{\text{obs}} \rightarrow B_i^{\text{obs}} \in \{\text{F},\text{P},\text{sil}\}^*$ and $S_j^{\text{son}} \rightarrow B_j^{\text{son}} \in \{\text{V},\text{A},\text{N}\}^*$ such that

$$B = B_1^{\text{obs}} B_1^{\text{son}} B_2^{\text{obs}} B_2^{\text{son}} \cdots B_{N^{\text{obs}}}^{\text{obs}} B_{N^{\text{son}}}^{\text{son}}.$$

(2) *Definition of sonorant intervocalic segments*: For each sonorant region, S_i^{son}, the contained vowel landmarks,

$$O_i^{\text{v}} = \{(t,f) \in O^{\text{v}} | t \in S_i^{\text{son}}\},$$

further partition the interval into a set of sonorant intervocalic regions. That is, if $N_i^{\text{v}} \equiv |O_i^{\text{v}}|$, then each interval S_i^{son} can be partitioned about the vowel landmarks it contains into $N_i^{\text{v}}+1$ sonorant intervocalic segments, denoted $\{S_{i0}^{\text{sc}}, \ldots, S_{iN_i^{\text{v}}}^{\text{sc}}\}$, where each S_{ij}^{sc} is an interval in time. A candidate sequence B is consistent with O^{v} if, for every $i \in \{1, \ldots, N^{\text{son}}\}$, the number of vowel ($V$) tokens in B_i^{son} is equal to N_i^{v}. Thus, for B consistent with O^{v}, we may write

$$B_i^{\text{son}} = B_{i0}^{\text{sc}} V B_{i1}^{\text{sc}} V \cdots V B_{iN_i^{\text{v}}}^{\text{sc}},$$

where the $B_{ij}^{\text{sc}} \in \{\text{A},\text{N}\}^*$ is the jth vowel-separated sonorant consonant sequence contained in B_i^{son}.

(3) *Model for* $P(B|O^{\text{seg}},O^{\text{v}})$: In general, the term $P(B|O^{\text{seg}},O^{\text{v}})$ can accommodate a probabilistic segmentation strategy. This amounts to considering candidate sequences B that may not be consistent with O^{seg} and O^{v} in the manner defined earlier. However, in this paper we consider only a hard segmentation. Thus, given the above-presented definitions and assuming independence of sonorant and obstruent regions according to the distinctive feature hierarchy, we may write

$$P(B|O^{\text{seg}},O^{\text{v}}) = \prod_{i=1}^{N^{\text{obs}}} P(B_i^{\text{obs}}) \prod_{j=1}^{N^{\text{son}}} P(B_j^{\text{son}})$$

if B is consistent with O^{seg} and equal to 0 otherwise. Furthermore, assuming independence of sonorant intervocalic regions, we may write

$$P(B_j^{\text{son}}) = \prod_{k=0}^{N_j^{\text{v}}} P(B_{jk}^{\text{sc}})$$

if B is consistent with O^{v} and equal to 0 otherwise. Putting it together, we have

$$P(B|O^{\text{seg}},O^{\text{v}}) = \prod_{i=1}^{N^{\text{obs}}} P(B_i^{\text{obs}}) \prod_{j=1}^{N^{\text{son}}} \prod_{k=0}^{N_j^{\text{v}}} P(B_{jk}^{\text{sc}}), \quad (2)$$

where we need only consider candidate sequences B that are consistent with the segmentation defined by O^{seg} and O^{v}.

(4) *Model for* $P(O^{\text{obs}},O^{\text{sc}}|B,O^{\text{seg}},O^{\text{v}})$: Since the broad class detectors at the leaf nodes of the hierarchy are dominated by the sonorant–obstruent distinction at the root node, we assume conditional independence of the form

$$P(O^{\text{sc}},O^{\text{obs}}|O^{\text{v}},O^{\text{seg}},B) = P(O^{\text{sc}}|O^{\text{v}},O^{\text{seg}},B) P(O^{\text{obs}}|O^{\text{v}},O^{\text{seg}},B).$$

Note that we only need to evaluate $P(O^{\text{sc}},O^{\text{obs}}|O^{\text{v}},O^{\text{seg}},B)$ for those B where $P(B|O^{\text{seg}},O^{\text{v}})$ is nonzero. Detector firings that lie in different segments are assumed to be independent as well. Therefore,

$$P(O^{\text{obs}}|O^{\text{v}},O^{\text{seg}},B) = \prod_{i=1}^{N^{\text{obs}}} P(O_i^{\text{obs}}|B_i^{\text{obs}})$$

and

$$P(O^{\text{sc}}|O^{\text{v}},O^{\text{seg}},B) = \prod_{i=1}^{N^{\text{son}}} P(O_i^{\text{sc}}|O^{\text{v}},B_i^{\text{son}}),$$

where $O_i^{\text{obs}} = \{(t,f) \in O^{\text{obs}} | t \in S_i^{\text{obs}}\}$ and $O_i^{\text{sc}} = \{(t,f) \in O^{\text{sc}} | t \in S_i^{\text{son}}\}$. Likewise, assuming independence between the firings in separate sonorant intervocalic regions, we may write

$$P(O_i^{\text{sc}}|O^{\text{v}},B_i^{\text{son}}) = \prod_{j=0}^{N_i^{\text{v}}} P(O_{ij}^{\text{sc}}|B_{ij}^{\text{sc}}),$$

where $O_{ij}^{\text{sc}} = \{(t,f) \in O_i^{\text{sc}} | t \in S_{ij}^{\text{sc}}\}$. Note that the preceding probability factorizations assume the pattern of detector firings in a given segment are generated solely by the corresponding subsequence of B. Collecting terms, we have

$$P(O^{\text{sc}},O^{\text{obs}}|O^{\text{v}},O^{\text{seg}},B) = \prod_{i=1}^{N^{\text{obs}}} P(O_i^{\text{obs}}|B_i^{\text{obs}}) \prod_{j=1}^{N^{\text{son}}} \prod_{k=0}^{N_j^{\text{v}}} P(O_{jk}^{\text{sc}}|B_{jk}^{\text{sc}}). \quad (3)$$

Given the form of Eqs. (2) and (3), we can recast the optimization problem of Eq. (1) as

$$B_{\text{opt}} = \arg\max_{B\in\Sigma^*} \prod_i P(O_i^{\text{obs}}|B_i^{\text{obs}})P(B_i^{\text{obs}}) \times \prod_{jk} P(O_{jk}^{\text{sc}}|B_{jk}^{\text{obs}})P(B_{jk}^{\text{sc}}) = \arg\max_{B\in\Sigma^*} \prod_{i=1}^{N^{\text{obs}}} P(B_i^{\text{obs}}|O_i^{\text{obs}}) \prod_{j=1}^{N^{\text{son}}} \prod_{k=0}^{N_j^{\text{v}}} P(B_{jk}^{\text{sc}}|O_{jk}^{\text{sc}}), \quad (4)$$

where we consider only B consistent with and O^{seg} and O^{v} as described earlier. Therefore, the global MAP optimization problem reduces to a set of segment-level optimizations which are performed independently. Furthermore, we assume that some fixed set of obstruent and sonorant intervocalic prior distributions generate all possible observations. We will present separate MAP models for the terms $P(B_i^{\text{obs}}|O_i^{\text{obs}})$ and $P(B_{jk}^{\text{sc}}|O_{jk}^{\text{sc}})$ in Sec. III C.

E. Connections to previous work

Many of the ideas going into our system may be traced to the work of Stevens and colleagues.[7] There are, however, significant challenges in translating the philosophy of distinctive features and acoustic landmarks into a viable computational strategy. In particular, coping with the demands of the immense variability in the speech signal makes it essential to find a statistical framework in which those ideas can be embedded.

The distinctive feature aspect of our approach is most closely related to the event-based system (EBS) developed by Juneja and Espy-Wilson,[36] which arrives at a broad class segmentation using a Viterbi-style decoding of frame-level distinctive feature probabilities. (This approach is also used as a component in the systems developed in Ref. 9.) Landmarks for further processing are derived from the transition points of this broad class segmentation. Therefore, the representation for each feature must run on a common frame rate to allow for the frame-level comparison, effectively making it a frame-based dynamic model.

Our approach deviates from EBS in two fundamental ways: (i) We immediately divide the utterance into a series of syllable-sized analysis units using a sonority profile and (ii) we immediately adopt a sparse point process representation composed of landmarks in time, which are probabilistically integrated to arrive at a broad class sequence. Since our broad class decoding procedure is not performed on the frame level, we have the freedom to vary the frame rate of the individual feature representations. Furthermore, the phonological dynamics are modeled entirely on the point process representation, completely distinguishing our approach from frame-based methods.

Still, others have deviated from frame-based approaches and created probabilistic landmark models. In particular, various versions of the SUMMIT system[10] model acoustic observations made on a phonetic segment level, as well as at landmarks coinciding with segment transitions. However, our approach distinguishes itself from SUMMIT in two important ways. First, since we employ a distinctive feature hierarchy, we reduce our recognition problem into independent syllable-sized chunks, minimizing the complexity of the model and exploiting the robust context dependence the distinctive feature hierarchy provides. Second, within syllable-sized chunks, we explicitly model the relative temporal dynamics of the landmarks, which in this syllable-centric setting are highly constrained.

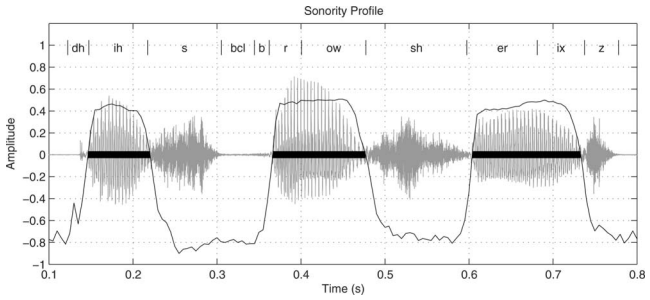


FIG. 6. A speech signal for the phrase "this brochure is" with the computed sonority profile overlaid. Sonorant segments are identified by horizontal bars.

III. TIMING TIER CONSTRUCTION AND INTEGRATION

Our timing tier representation requires the construction of a sonority segmenter and six broad class landmark detectors. The resulting landmark set must then be probabilistically integrated to perform a phonological decoding. In the following, we outline the theoretical and algorithmic details involved with our computational approach to these problems. Experimental details and results will be covered separately in Sec. IV.

A. Sonority segmentation

Computationally, a sonority segmentation may be accomplished using any available machine learning method. In our implementation, we employ support vector machines (SVMs). This popular machine learning technique involves solving the optimization problem

$$f^* = \min_{f\in\mathcal{H}_K} C\sum_{i=1}^{l} \max(0, 1 - y_i f(x_i)) + \|f\|_K^2,$$

for a decision surface f^* restricted to a representing kernel Hilbert space (RKHS) $\mathcal{H}_K$ for some kernel function K. Here, $\{(x_i, y_i)\}_{i=1}^l$ are the labeled training data and $\|f\|_K$ indicates the RKHS norm. Using the method of Lagrange multipliers, this convex optimization problem can be solved using a quadratic programming solver.

The SVM hinge-loss weight parameter C must be chosen appropriately regardless of the kernel used. Furthermore, we employ the radial basis function (RBF) kernel, $K(x,y) = \exp(-\gamma\|x-y\|^2)$, which introduces a second parameter, γ, the Gaussian argument coefficient. These parameters are chosen using holdout validation with the training set data. We employ a 39-dimensional mel frequency cepstral coefficient feature set.[37] However, any reasonable feature set may be used here, including possibly more robust acoustic parameters.[31–33,38,39]

Once the SVM is trained, to determine the segmentation we simply threshold the sonority SVM output, as shown in Fig. 6. In the subsequent stages of the system, *sonorant* re-

gions will refer to the segments above this threshold, while *obstruent* regions will refer the segments that fall below. Ideally, sonorant regions will contain sequences of vowels, approximants, and nasals; obstruent regions will contain sequences of silences, stops, and fricatives. In practice, the sonority segmentation is imperfect, and incorrect broad class content can be present in a particular region.

B. Broad class landmark detectors

The construction of detectors involves first training appropriate SVM classifiers for each broad class in Table I and then converting their real output into sparse sets of landmarks.[40]

1. Constructing the classifiers

If our sonority segmenter performed without error, there would be no presence of sonorant phonemes in predicted obstruent regions, and vice versa. In this ideal setting, the nasal classifier, for example, would only need to be constructed to differentiate between nasals and other sonorant classes. However, with an imperfect sonority segmentation, the classifiers need to be proficient at discriminating against all other possible phonemic content. For this reason, each one-versus-all classifier is trained with examples across the entire phonetic space, not just those in its sonority class. However, when integrating detectors, we only consider the firings of broad class detectors that are consistent with the initial segmentation. Fortunately, the robustness of the sonorant–obstruent distinction results in small differences between the performance of SVMs trained within their sonority class versus those trained across all phones. In the language of SVMs, the vast majority of the support vectors discriminate within the given sonority class. Our broad class SVMs also employ the RBF kernel, so we must again determine optimal C and γ parameters for each classifier via hold-out validation.

Since each classifier processes the signal independently, their construction can be specialized according to the individual broad class content. While we choose 39-dimensional MFCC features for the silence, fricative, nasal, approximant, and vowel SVMs, the framing parameters and frequency ranges used for each vary. Furthermore, our stop classifier employs energy and Wiener entropy parameters shown to be successful in this setting.[29] The modularity of this independent detector approach provides maximal flexibility for future development of our framework.

2. From classifier to detector

The output of each SVM is a real number for each frame of the signal. In general, after thresholding this series, we define the landmark time as the position of any local maximum of the SVM output and the landmark strength as the corresponding maximal values. The one exception made to this landmark picking strategy is for the vowel detector. It is common for the output of the broad class classifiers to experience multiple local maxima within a single phone as a result of acoustic variation arising from coarticulation. This does not pose a problem when landmarks are subsequently processed by our probabilistic integration scheme, as resulting degenerate (multiple correct within a single phone) detections are accommodated (see Sec. III C). However, in the positioning of vowel landmarks, degenerate detections will necessarily result in vowel insertions.

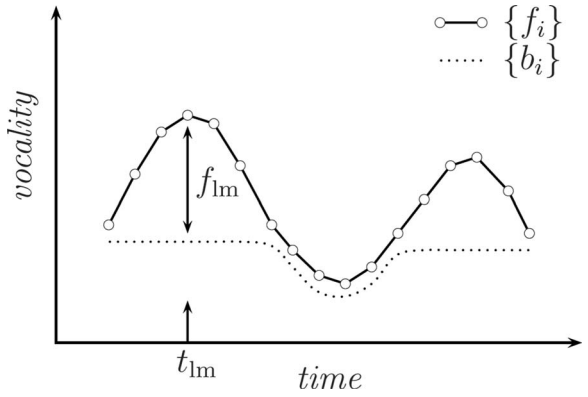


FIG. 7. A schematic plot of a series $\{f_i\}$ with its corresponding dynamic baseline $\{b_i\}$ for the given iteration. The corresponding landmark time t_{lm} and strength f_{lm} for this iteration are also shown.

To address this complication, we require a more sophisticated approach to choosing vowel landmarks given the continuous output of the vowel classifier. For this, we employ an adaptation of the "convex-hull" approach presented in Ref. 39 to recursively compute a time-dependent baseline. The input of the algorithm is the vowel classifier output time series, $\{f_i\}_{i=1}^{N}$, and proceeds as follows (see Fig. 7):

(1) Let $k=\arg\min_i\{f_i\}$ be the index of the absolute minimum of $\{f_i\}$. We can then compute the dynamic baseline, $\{b_i\}$,

$$b_i=\begin{cases}\min\{b_{i-1},f_i\} & \text{for } i=2,\ldots,k-1\\ \min\{b_{i+1},f_i\} & \text{for } i=k+1,\ldots,N-1\\ f_i & \text{for } \quad i=1,k,N.\end{cases}$$

This baseline is monotonically decreasing to the absolute minimum and monotonically increasing afterward.

(2) Create a new series $d_i=f_i-b_i$, equal to the difference between the original series and the dynamic baseline.

(3) Define a new landmark with $t_{\mathrm{lm}}=\arg\max_i\{d_i\}$ and $f_{\mathrm{lm}}=\max_i\{d_i\}$. Split $\{d_i\}$ about t_{lm} into two series, $\{l_i\}$ and $\{r_i\}$.

(4) Repeat all steps for both $\{f_i\}=\{l_i\}$ and $\{f_i\}=\{r_i\}$ while preserving absolute time positions, until a recursion depth is reached that overgenerates the number of candidate landmarks. This candidate set is then pruned by thresholding on the values of f_{lm}. This threshold is chosen empirically on a validation set.

Using this scheme, the strength of the landmark is determined relative to the baseline computed for the given iteration. While the amplitude of the local maxima may be large, nearby candidates compete with respect to a baseline computed in the local region. Therefore, small variations of the detector output that would otherwise result in degenerate landmarks are rejected with an appropriate choice of threshold on the difference series, $\{d_i\}$.

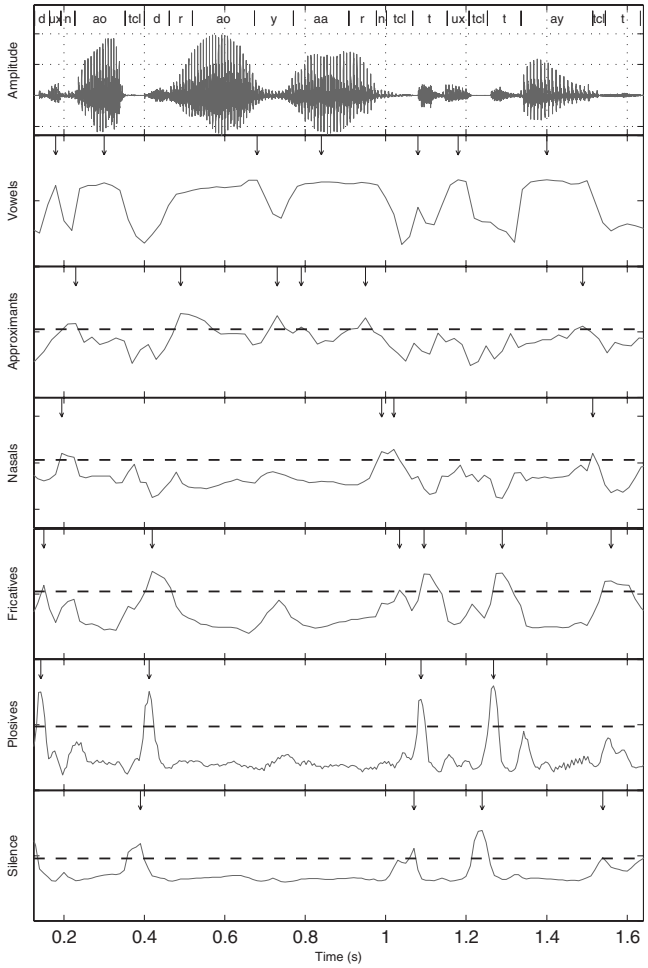


FIG. 8. The signal, broad class detector outputs and thresholds (dashed lines), and resulting landmarks (arrows) for the sentence "Do not draw yarn too tight."

Figure 8 shows the speech signal for the sentence "Do not draw yarn too tight," along with the output of the six classifiers, the thresholds used,[41] and the corresponding landmark positions. In Fig. 8, we see the vowel classifier reaches a local maximum at about 1.5 s. However, since it is lower than the adjacent local maximum at 1.4 s, the dynamic baseline algorithm prevents the insertion of a degenerate landmark within the /ay/ phoneme.

C. Probabilistic landmark integration

We now return to the task of probabilistically integrating the landmarks within each obstruent and sonorant intervocalic segment. This is akin to modeling $P(B_i^{\text{obs}}|O_i^{\text{obs}})$ and $P(B_{jk}^{\text{sc}}|O_{jk}^{\text{sc}})$ of Eq. (4) (see Sec. II D). To simplify the following discussion, we reset the above-used global notation, where (B,O) now refers to either $(B_i^{\text{obs}},O_i^{\text{obs}})$ or $(B_{jk}^{\text{sc}},O_{jk}^{\text{sc}})$, depending on the segment type we are discussing.

Now, if each broad class detector operated flawlessly, we could simply chronologically sort the detections in each sonorant or obstruent region, resulting in a perfect transcription of the content. However, in the presence of false positives and negatives, modest individual detector mistakes can combine to drastically reduce performance. For example, if each detector operated at an admirable 10% precision-recall equal error rate, this naive method of decoding would have a combined 50% insertion rate. Clearly, a more sophisticated integration strategy is required to clean up misfires.

To address the complication of spurious and missed landmarks, we have developed a framework for probabilistic segment decoding (PSD) based on a MAP estimate of the broad class landmark sequence in a given obstruent or sonorant intervocalic region. To see how this works, consider an interval of the speech signal (T_1,T_2) of duration $T=T_2-T_1$ that when combined with the activity of N broad class detectors, defines a set of observables

$$O=\{T,O_{X_1},\ldots,O_{X_N}\},$$

where each O_{X_i} denotes the observables for the class X_i detector. These consist of L_{X_i} time–strength paris (one per detection) which we denote as

$$O_{X_i}=\{(t_1^{X_i},f_1^{X_i}),\ldots,(t_{L_{X_i}}^{X_i},f_{L_{X_i}}^{X_i})\},$$

where we have converted the absolute landmark times to the fraction of the segment that passes before the landmark occurs. That is, if t is an absolute landmark time, the corresponding observable is $t^{X_i}=(t-T_1)/T$.

At this point we can immediately write down a simple MAP estimate of the segment broad class sequence, $B_{\text{opt}}=\max_B P(B|O)$. However, in the context of our hierarchical landmark-based system, we would like our model to also estimate which landmarks within the region were correct and which were misfires. With this information, we can proceed with transcription refinement at true landmarks. To address this, we can define a set of indicator variables,

$$H=\{H_{X_1},\ldots,H_{X_N}\},\quad H_{X_i}=\{h_1^{X_i},\ldots,h_{L_{X_i}}^{X_i}\},$$

where $h_k^{X_i}=1$ if the kth detection of class X_i is a true positive, and 0 otherwise.

Our goal is to determine the most likely broad class sequence, B_{opt}, and which landmarks construct it, H_{opt}. Given the above-presented nomenclature, this can be accomplished by computing the MAP estimate,

$$\begin{aligned}(B_{\text{opt}},H_{\text{opt}}) &= \arg\max_{B,H} P(B,H|O)\\ &= \arg\max_{B,H} P(O|B,H)P(H|B)P(B).\end{aligned}\tag{5}$$

Notice each probability term in this optimization problem is estimable by application of the sonority segmentation and broad class detectors to a corpus of transcribed training data. Furthermore, since our approach is to first segment the utterance down to short analysis units consisting of a limited number of phonemes, we can accomplish optimization by simply calculating the likelihood for all possibilities. If we attempted the same exhaustive approach for word or sentence-long reconstruction units, this combinatorial problem would become prohibitively cumbersome.

1. Decoding obstruent regions

Given an obstruent region determined by the sonority segmentation and the set of observables determined by the obstruent landmark detectors, it remains to apply our proba-

bilistic segment decoding algorithm introduced earlier. [This is equivalent to modeling $P(B_i^{\text{obs}}|O_i^{\text{obs}})$ of Eq. (4) in Sec. II D.] The goal is to determine a transcription of stops, fricatives, and silences present in each obstruent region (i.e., $B \in \{\text{P},\text{F},\text{sil}\}^*$). Note that silence landmarks are included in both the provided set of observables and target sequence search space. However, their inclusion in the final broad class transcription is not necessary for typical applications, and thus will be ignored in our performance evaluations.

Specializing for the task of obstruent segment decoding, we can further simplify the general MAP estimation problem of Eq. (5) with several independence assumptions (in the following expressions, $\mathcal{C}_{\text{obs}}=\{\text{sil},\text{P},\text{F}\}$):

(1) The behaviors of the broad class detectors are independent of each other and the obstruent segment duration,

$$P(O|B,H) = P(T|B) \prod_{X \in \mathcal{C}_{\text{obs}}} P(O_X|B,H).$$

(2) The behavior of each broad class detector depends only on its own indicator variables and not those of other detectors,

$$P(O_X|B,H) = P(O_X|B,H_X).$$

(3) The detection correctness pattern for one broad class detector is independent of that of the others,

$$P(H|B) = \prod_{X \in \mathcal{C}_{\text{obs}}} P(H_X|B).$$

(4) The times of the detections for a particular class are independent of each other. That is, detection times depend only on the broad class sequence encountered and whether the detection is a true positive,

$$P(t_1^X, \ldots, t_{L_X}^X|B,H_X) = \prod_{i=1}^{L_X} P(t_i^X|B,h_i^X).$$

(5) The strengths of the detections for a particular class are independent of each other and the broad class sequence encountered. That is, detection strengths depend only on whether the detection is a true positive,

$$P(f_1^X, \ldots, f_{L_X}^X|B,H_X) = \prod_{i=1}^{L_X} P(f_i^X|h_i^X).$$

(6) Detector strengths and times are independent,

$$P(O_X|B,H_X) = \prod_{i=1}^{L_X} P(t_i^X|B,h_i^X)P(f_i^X|h_i^X).$$

While the extent of the validity of these assumptions has not been rigorously established, their inclusion in the formulation vastly reduces the number of training sentences required to estimate the component distributions. Under these independence assumptions, the optimization problem of Eq. (5) reduces to

$$(B_{\text{opt}},H_{\text{opt}}) = \arg\max_{B,H} P(T|B)P(B) \prod_{X \in \mathcal{C}_{\text{obs}}} P(H_X|B) \times \prod_{i=1}^{L_X} P(t_i^X|B,h_i^X)P(f_i^X|h_i^X). \quad (6)$$

2. Decoding sonorant regions

The decoding of sonorant regions involves a significant complication over the obstruent task. Phonotactic constraints of English and the limitations of the human vocal apparatus make the production of long obstruent sequences extremely unlikely. In fact, in TIMIT all obstruent sequences have a length of four phones or less. This is not the case for sonorant regions; for example, the sentences "We were away all year." and "When will you hear me?" are entirely sonorant. Therefore, in order to limit the combinatorial complexity of sonorant region decoding, we must further segment the signal into smaller, more easily analyzable units. Here, the natural choice is to again turn to the syllable to set the meter of analysis. Since syllables are tied to the vowels present in the sonorant regions, the logical points of separation are vowel landmarks produced by the vowel landmark detector.

Within a given sonorant region, L vowel landmarks determine a series of $L+1$ intervocalic regions that ideally contain sequences of approximants and nasals. Note that the first and last of these regions are bounded by adjacent obstruent regions as determined by the sonority segmentation.[42] As done for obstruent regions, we can apply the probabilistic segment decoding approach to reconstruct the approximant and nasal content of each intervocalic region. [This is equivalent to modeling $P(B_{jk}^{\text{sc}}|O_{jk}^{\text{sc}})$ of Eq. (4) in Sec. II D.] Now, in the case of obstruent regions, T was simply the duration of the entire segment. For intervocalic regions, T is defined as the time elapsed either between adjacent vowel landmarks, between a landmark and an adjacent sonorant region boundary, or, if there are no vowel landmarks, the entire length of the sonorant region. The nasal and approximant landmarks (time and strength) round out the set of observables.

We employ the same set of observable independence assumptions for intervocalic decoding as listed in Sec. III C 1 for obstruent regions. The optimization problem again takes the form of Eq. (6), where the obstruent class set $\mathcal{C}_{\text{obs}}$ is replaced with $\mathcal{C}_{\text{iv}}=\{\text{A},\text{N}\}$ and the possible sequences are $B \in \{\text{A},\text{N}\}^*$.

3. Estimating the probability distributions

Our probabilistic framework, under the independence assumptions described earlier, requires the measurement of several prior distributions. These include:

(1) $P(B)$ for each possible intervocalic and obstruent segment broad class sequence, B.
(2) $P(T|B)$ for each possible broad class sequence, B.
(3) $P(H_X|B)$ for each possible broad class sequence/detector pair, (B,X).

(4) $P(t^X|B,h)$ for each possible broad class sequence/detector/indicator variable value triplet, (B,X,h) for $h \in \{0,1\}$.
(5) $P(f^X|h)$ for each broad class detector/indicator variable value pair, (X,h) for $h \in \{0,1\}$.

Given segmented data, i.e., speech wave forms along with their transcriptions, estimating each of these distributions is fairly straightforward. Thus, $P(B)$, is simply a distribution on a finite set of sequences that occur in obstruent and intervocalic regions, respectively. For each segment S, if we knew the values of H, B, O_{seg}, O_V, O_{sc}, and O_{obs} (see Sec. II D), then we could estimate $P(T|B)$, $P(t^X|B,h)$, and $P(f^X|h)$ as distributions on the real line.

The first step in this process involves running the sonority segmenter on each training sentence. This will result in several sonorant and obstruent regions per sentence for analysis and eventual contribution to the distributions. Likewise, vowel landmarks are computed for the sonorant regions, resulting in multiple intervocalic regions per sonorant region for contribution to the distribution estimation. In general, we may arrive at the "true" value of B for each segment by force aligning the phonetic transcription with the sequence of obstruent and sonorant intervocalic segments determined by the sonority segmenter and vowel landmarks. Since our segmentation may insert or delete segments, we discard any elements of each segment's B that are inconsistent with the segment type. Thus, we ensure the true B is always consistent with the segmentation.

We set T to be the measured duration according to the segmentation and not the actual transcription interval. The detectors result in a collection of landmarks, each consisting of a time–strength pair, (t_i^X, f_i^X). The corresponding indicator variables, $\{h_i^X\}$, can be determined by checking time position against the transcription. This indicator value will determine whether that time–strength pair will be logged in the $h=0$ or $h=1$ distributions. In addition, the individual indicator variables will be combined to form sequences $H_X=(h_1^X,\ldots,h_{L_X}^X)$ for the $P(H_X|B)$ distributions.

There are several possible ways of handling the estimation of these distributions. The most computationally straightforward is the histogram method, which involves simply maintaining a list of all values encountered for each. Using this list, a probability $P(X|Y)$ is calculated by

$$P(X|Y) = \frac{\text{No. cases of } X \text{ in } Y \text{ list}}{\text{length of } Y \text{ list}}.$$

For the discrete distributions, $P(B)$ and $P(H_X|B)$, this histogram prescription is adequate. However, for scalar variable distributions in f, t, and T, we instead implement uniform (i.e., rectangular) kernel density estimation, for which we much choose an appropriate kernel width. This leads to the introduction of three kernel width parameters into the model: Δf, Δt, and ΔT. For example, given a list for the distribution $P(f^X|h=1)$, the probability at a given strength f_0 is estimated by

$$P(f_0|h=1) = \frac{\text{No. cases of } f^X \in [f_0-\Delta f, f_0+\Delta f] \text{ in } h=1 \text{ list}}{\Delta f \times \text{length of } h=1 \text{ list}}.$$

In practice, to limit the model parameters, we choose one set of kernel widths for obstruent region decoding and one set for intervocalic region decoding. However, separate sets could be assigned for the observables for each broad class.

These approaches require a significant amount of training data to provide good distribution estimates. To circumvent this problem, more sophisticated techniques of distribution estimation may also be used. We tried applying Gaussian mixture models, but they resulted in inferior performance. For example, parametric modeling or nonuniform kernel smoothing may improve estimates in the face of limited examples. Exploring these methods lies outside the scope of this paper.

IV. EXPERIMENTAL DETAILS AND RESULTS

In the following, we present the performance results for each of the components of our landmark-based recognition system, as well as the end-to-end broad class transcription performance. When applicable, we evaluate a given component assuming ideal performance of the other elements of the system. This allows us to isolate the merits and shortcomings of each component to aid in future development of integration strategies.

A. Sonority segmentation performance

The support vector machine for the sonority segmenter was trained on a total of 100 "sx" (phonetically compact) and "si" (phonetically diverse) sentences chosen at random from the training section of the TIMIT database. For each of these sentences, 39-dimensional mel frequency (40 spectral bands) cepstral coefficients spanning the full frequency range (0–8 kHz) were computed in 10 ms windows every 5 ms. For the 100 sentences, this translates into approximately 60 000 39-dimensional training vectors, nearly evenly distributed between sonorant and obstruent regions. We employed the SVM^{light} software package[43] to construct the SVM. An operating threshold of 0.1 and the SVM parameters $C=0.0816$ and $\gamma=5\times10^{-4}$ (see Sec. III A) were chosen via holdout validation. This resulted in a frame-level training error of 6.12%.

Likewise, we tested the sonority segmentation performance on 100 randomly chosen sx/i test sentences. Evaluating performance at this initial stage of the overall system is not entirely straightforward. The frame-level test error of 6.44% provides a useful first approximation of SVM performance, but we are primarily interested in evaluating the performance in the context of our overall system architecture. Since our initial sonority segmentation is rigid, we must get this initial classification correct on a phone level to have a chance of correct transcription at later stages. (We will return to this shortcoming in Sec. V.)

This argument points to two phone-level performance metrics:

TABLE II. Performance of the sonority segmentation.

F_{min}	C_{son} (%)	C_{obs} (%)
0.10	98.5	95.4
0.33	96.6	92.9
0.50	95.0	89.3
0.67	93.4	85.8
0.90	82.1	68.7

(1) C_{son}=percentage of the individual sonorant phones for which at least a fraction F_{min} of its duration falls in a single sonorant segment as determined by the sonority segmenter.
(2) C_{obs}=percentage of the individual obstruent phones for which at least a fraction F_{min} of its duration falls in a single obstruent segment.

Since the SVM output is effectively smoothed by the windowing parameters, a perfect segmentation down to the sampling interval (1/16 000 s) is impossible. Inaccurate TIMIT transcription time boundaries further complicate the matter. Therefore, setting the minimum overlap F_{min} to one or even close to one is unreasonable. Further, since the entire phone need not be present in a given region for successful decoding, segmentation performance for low F_{min} values can still be a good indicator for success in later stages. It is also important to note that segment insertions do not necessarily preclude correct decoding at later stages, as spurious sonorant or obstruent regions may be decoded to be empty. However, each phone must have positive overlap with a proper sonority segment to have a chance to be decoded.

The values of the performance measures as evaluated on our test set are summarized in Table II for various minimum overlap requirements. As expected, the correctness percentages drop as we require larger fractional overlaps. However, the rate of performance decline is higher for obstruent phones. This is largely a result of their shorter average duration, for which systematic errors caused by the windowing parameters and transcription inaccuracies constitute a larger relative portion.

Figure 9 shows the phonetic breakdown of the segmentation errors. More precisely, it displays the number of instances that the durational majority of each phone was placed into an incorrect segment; the sonorant and obstruent phone errors correspond to the 5.0% and 10.7% error rates of the F_{min}=0.5 line of Table II, respectively. For sonorant phones, the largest error contributors are the approximant [hh] and the nasals [m] and [n]. The phone [hh] is a glottal transition whose sonority status is not always well defined (varying definitions may be found in the literature). Therefore, given arbitrary context, mistakes either way are to be expected. Similarly, the nasals are closest to the sonorant–obstruent boundary, so it is not surprising that mistakes occur.

The most prominent mistakes for obstruent phones are for the glottal stop [q] (allophone of /t/) and flap [dx] (allophone of /d/ or /t/). Both allophones are typically of extremely short duration and are surrounded by vowels. Therefore, the sonority profile tends to peak for a minority of the duration or not at all, resulting in a significantly diminished chance of their successful decoding later on. Unfortunately, these allophones are somewhat common. From the error breakdown, we also find that voiced fricatives and stops tend to contribute a higher error rate than their unvoiced counterparts of the same place (e.g. [d] vs [t] or [v] vs [f]). Again, this is not surprising as sonority is largely a measure of the periodicity introduced by glottal voicing and so the SVM training follows suit.

B. Landmark detector performance

Constructing the six landmark detectors required the construction of six support vector machines trained to recognize phones of the target class. We worked with a set of 100 randomly chosen TIMIT sx/i training sentences, though not all frames of all sentences were used for every detector. With full flexibility in constructing each detector, we worked with several representations. For the vowel, approximant, nasal, fricative, and silence detectors, we use 39-dimensional mel frequency (40 bands) cepstral coefficients, but the window size (T_{win}), step size (T_{step}), and frequency range (F_{range}) parameters varied according to Table III. The SVM C and γ

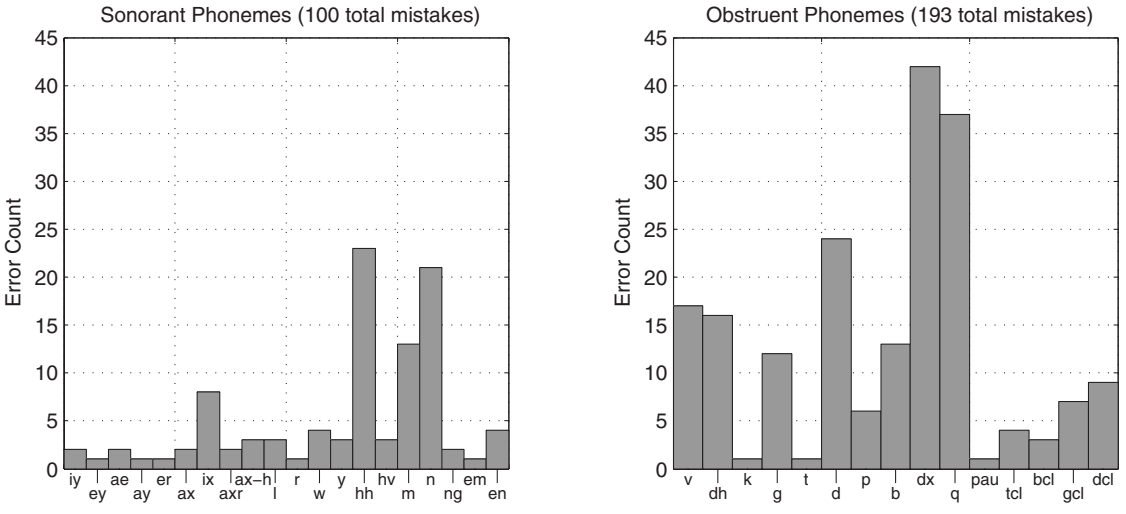


FIG. 9. Phonetic composition of sonority segmentation errors for F_{min}=0.5. Phones for which no mistakes were made are not shown.

TABLE III. Representation and training parameters for the landmark detectors.

Detector	T_{win} (ms)	T_{step} (ms)	F_{range} (kHz)	C	γ	E_{train} (%)
Vowel	40	20	0–4	0.600	0.000 65	10.3
Approximant	20	20	0–8	0.632	0.000 56	19.2
Nasal	30	15	0–8	0.469	0.000 80	6.0
Fricative	30	15	0–8	0.236	0.000 71	6.9
Stop	35	5	N/A	0.064	0.016	6.2
Silence	20	10	0–8	0.017	0.000 53	6.0

parameters, both found via holdout validation, as well as the frame-level training errors (E_{train}) are also listed. For training these five SVMs, all frames centered within the desired phoneme boundaries were considered positive examples.

For the stop detector, we used the acoustic parameter prescription provided by Ref. 29 as an alternative to the MFCC representation, though we modified the frame rate to reduce computational costs. There are three quantities involved in this prescription: total energy, energy above 3 kHz, and the Wiener entropy, $\Sigma_i \log S_i - \log(\Sigma_i S_i)$, where $\{S_i\}$ is the discrete Fourier amplitude spectrum. Each of these parameters was computed in nonoverlapping 5 ms windows. Each 21-dimensional training vector consisted of seven consecutive frames of this type, spanning an effective window width of 35 ms. Given this representation, the SVM has implicit access to the differences in the three parameters over the seven component frames, allowing for the identification of inflection points associated with the stop–burst transition. Furthermore, the 5 ms temporal resolution provides adequate precision for the detection of this transition. The training setup and parameters for the stop detector are also provided in Table III. For training, only the frame centered closest to the stop–burst transition was considered a positive example. Therefore, there were only as many positive examples per sentence as there were stop phones present. For this reason, we limited the number of negative frames to a random sample that is five times the number of positive frames.

Figure 10 shows the phoneme-level landmark performance curves for each detector as a function of threshold. Here, the false positive (FP) rate is defined as the number of spurious detections divided by the number of negative (i.e., not target) phones; the false negative (FN) rate is defined as the number of missed detections of the target class divided by the number of phones of the target class present. Since these rates are computed on a per landmark basis, arbitrarily decreasing the threshold does not necessarily mean the false negative rate goes to zero, distinguishing the curves from the traditional receiver operating characteristic (ROC) variety. However, the operating thresholds are chosen at the point of equal FN and FP rates, as overlaid on each plot. For the vowel detector curve, we include degenerate (extra correct detections within a single phone) vowel detections in the FP count, as they will necessarily result in insertions. However, for the other five detectors, degenerate landmark detections are ignored, as the probabilistic integration model can accommodate such mistakes.

Unfortunately, the stop detector prescription we use was not designed to identify the above-mentioned allophones [q] and [dx]. However, for our purposes we cannot ignore these somewhat common sounds, so our detector phoneme-level error rate is seven points higher than the performance quoted in Ref. 29. Ideally, these allophones would require the construction of their own detector, which can either function on its own or as part of a composite stop detector. We will return to this issue in our discussion of future research directions in Sec. V.

Table IV shows the broad class breakdown of insertion errors made by each detector, along with the number of correct, degenerate, and deleted landmarks. There are several points to note from this breakdown. First, we find that a significant majority of errors are made between broad classes

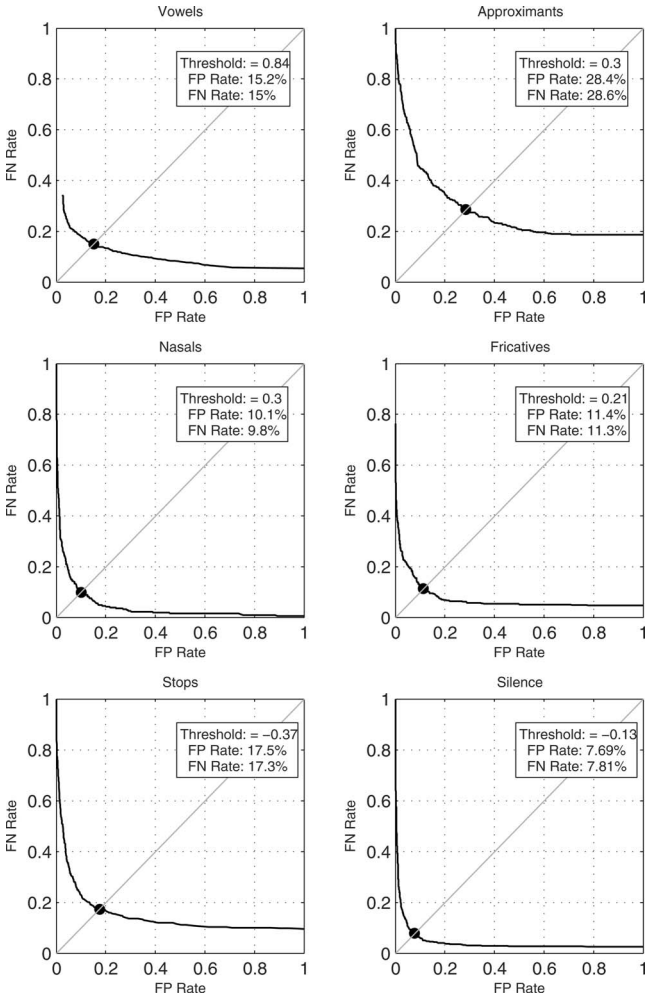


FIG. 10. Phoneme-level ROC curves for the six detectors. The operating thresholds taken at the equal error points are listed on the plots.

TABLE IV. Insertion composition by broad class, number correct, number deleted, and number of degenerate detections for each landmark detector.

Detector	V	A	N	F	P	sil	Degen	Del
V	1036	84	26	96	112	62	21	174
A	606	329	159	28	103	30	51	122
N	125	82	276	42	32	72	81	30
F	7	8	12	400	229	128	237	50
P	160	93	47	182	473	74	63	98
sil	28	18	46	60	24	722	764	42

of the same sonority superclass. For example, only 7% of the fricative detector misfires occur within sonorant phones. This trend validates our choice of the sonorant feature as an appropriate initial point of speech segmentation. Still, we find that mistakes across the sonorant–obstruent distinction do occur. However, so long as the sonority segmentation is correct in the region, such spurious landmarks will be discarded.

Next, we find that the vowel detector results in the lowest degenerate landmark rate (only 21 degenerate vowel landmarks in 1210 vowel phones). This is a direct result of the dynamic baseline algorithm presented in Sec. III B 2. While using a fixed threshold would result in a slightly higher detection rate, the massive resulting increase in degenerate detections would greatly decrease the accuracy. However, the higher detection rate that a fixed threshold provides is beneficial for the other detectors, since the landmark integration strategy is not significantly harmed by degeneracy. As we can see from Table IV, degeneracy rates can be quite high for the longer fricative phones and silence regions.

The weakest link by far in the set is the approximant landmark detector. While the intervocalic decoding model can clean up these mistakes to some extent, poor detector performance definitely translates into unstructured probability distributions and ultimately poor integration performance. However, the low relative approximant incidence rate in typical speech still allows for good overall performance. We will return to the topic of detector-level inadequacies in Sec. V.

C. Obstruent region decoding performance

To separate the performance of the obstruent decoding model from that of the sonority segmenter, we conducted experiments using the actual sonority segmentations provided by the TIMIT transcription for both training and testing.

Prior distribution data were collected from 1000 randomly chosen sx/i training sentences. According to the prior $P(B)$ estimated from TIMIT training data, there are 42 possible sequences of stops, silences, fricatives that may lie in any obstruent region. Using an additional 100 training sentences, we determined optimal-accuracy kernel width parameters of $\Delta f=0.03$, $\Delta t=0.03$ (i.e., 3% of the obstruent region), and $\Delta T=10$ ms. Note that optimal kernel width parameters differ when estimated sonority segmentations are used.

We tested on all 1344 sx/i sentences contained in the TIMIT test set. For each obstruent region, the decoding results in a predicted transcription. We evaluate this prediction relative to the actual sequence present using minimum string edit distance alignment. For N phones encountered, this evaluation will result in the number of correctly identified phones (C), number of deleted phones (D), number of substituted phones (R), and number inserted phones (I). The accuracy can then be computed as $A=1-(D+R+I)/N=(C-I)/N$.

Table V shows the transcription performance for several variations of the decode procedure. The first is a naive measure of baseline performance without using the probabilistic model. Here, we simply chronologically sort the landmarks above the appropriate operating threshold in each obstruent region. The predicted sequence is simply the corresponding broad classes of these landmarks. The second method is the standard implementation of the probabilistic decoding method outlined in this paper. Finally, the two "Rank$\leqslant N$" methods assume we have an oracle that can identify the true obstruent region sequence if it is one of the N most probable sequences. (The standard decode is equivalent to "Rank$\leqslant 1$.") In all of these variations, we ignore silence landmarks in the performance analysis since their presence is not necessary in the final transcription.

The poor naive baseline performance illustrates the main problem with integrating multiple error prone detectors: correctness rates average, but insertion rates add together. However, our probabilistic decoding approach effectively cleans up false detections, admitting an insertion rate of only 6% while maintaining the correctness rate of the baseline. As we consider more than just the top sequence, prediction accuracy quickly improves further. These Rank$\leqslant N$ methods provide a ceiling estimate of expected model performance when higher-level linguistic constraints are incorporated, such as a language model. In such a setting, multiple high-probability predictions can be considered in the context of a word or phrase, effectively providing an approximation to an oracle function. While our model attains a 77% phoneme recognition accuracy in the top choice, it is exceedingly good at

TABLE V. Obstruent region decoding performance on 17 525 phones (fricative and stops) contained in 15 766 obstruent regions.

Method	Accuracy	% correct	% ins	% del	% repl
Baseline	42.0	79.2	37.2	14.9	5.9
Std. deccode	77.0	83.0	6.0	6.0	11.1
Rank$\leqslant 2$	89.2	92.1	2.9	3.8	4.2
Rank$\leqslant 3$	93.8	94.7	0.9	3.0	2.4

TABLE VI. Intervocalic region decoding performance in 12 915 phones (approximants and nasals) contained in 36 255 intervocalic regions.

Method	Accuracy	% correct	% ins	% del	% repl
Baseline	25.5	54.0	28.5	42.5	3.5
Std. decode	53.0	69.9	16.9	23.2	6.9
Rank $\leqslant 2$	85.1	90.4	5.3	6.7	2.9
Rank $\leqslant 3$	95.1	96.8	1.7	2.1	1.0

paring down the original 42 possibilities to a few candidate sequences, with close to 90% accuracy in the top two ranking candidates alone.

D. Intervocalic region decoding performance

For the performance evaluation of intervocalic decoding, we again determine region boundaries using the actual TIMIT transcription for both training and testing. The transcription center point of each vowel is used in place of detected vowel landmarks. This procedure allows us to isolate the performance of the intervocalic decoding from that of both the sonority segmenter and the vowel landmark detector. We estimate the prior distributions using the same 1000 randomly chosen TIMIT training sentences used for the obstruent region model. We again accomplish alignment using minimum string edit distance and use holdout validation to estimate the optimal-accuracy kernel width parameters of $\Delta f = 0.1$, $\Delta t = 0.1$ (i.e., 10% of the intervocalic region), and $\Delta T = 10$ ms.

We again test on all 1344 sentences contained in the TIMIT test set. According to the prior $P(B)$ estimated from the TIMIT training data, there are now only 12 possible sequences of nasals and approximants that may lie in any intervocalic region (including the empty sequence). Table VI shows the transcription performance for the same methods studied for obstruent region decoding. We find significantly lower performance of both the naive baseline and the standard decode relative to obstruent region decoding. This is largely due to the exceptionally poor performance of the approximant detector, as discussed in Sec. IV B. In particular, the high insertion rate of approximant landmarks in nasal phones is especially detrimental to the reliability of probabilistic predictions. Still, our decoding method significantly reduces the insertion rate while increasing the correctness percentage over the baseline performance, resulting in more than twice the accuracy. The ranking methods result in an even more striking performance improvement. When considering just the two most likely sequences, the accuracy improves 32% (absolute) over the standard decoding method. This portends great improvements in this module of the system when higher-level linguistic constraints are imposed.

E. Overall performance

We now turn to the overall performance of our landmark-based broad class recognizer, implementing the sonority segmenter, landmark detectors, and probabilistic segment decoding for both obstruent and intervocalic regions. Because we are now using estimated sonority boundaries and vowel landmarks, the optimal kernel width parameters have been retuned using holdout validation, as listed in Table VII. We find that the intervocalic parameters and the obstruent region Δf parameter are the same as when we used the actual TIMIT transcription boundaries. However, optimal values for the two timing parameters (Δt and ΔT) for obstruent regions have increased. This is largely a result of an imprecise sonority segmentation, resulting in a significant smear of the probability distributions. In this degraded setting, the model performs better when the distributions are more highly smoothed.

TABLE VII. Optimal kernel width parameters for probability distribution estimation.

Region type	Δt	ΔT (ms)	Δf
Obstruent	0.2	25	0.03
Interocalic	0.1	10	0.1

We tested our system and four continuous CMU Sphinx-3 HMM variations,[44] using both context-independent (CI) and dependent (CD) decoding with either broad class (BC) or individual phone (Ph) three-state models. Each HMM was trained on all 3696 TIMIT sx/i training sentences, using standard 39-dimensional mel frequency cepstral coefficients (Sphinx feature set 1s_c_d_dd), eight-mixture Gaussian observation densities, no skip transition. Furthermore, in our HMM experiments, no language model is applied (including no transition probability rescaling), restricting the study to the domain of pure speech recognition.

In our system, probabilistic segment decoding is a context-dependent approach, though the sonority segmentation and vowel landmark detection methods are context independent. Therefore, our composite system is only partially context dependent. The context-dependent HMM system uses triphone models in all cases, so its use of contextual information is more complete. The complexity of our system is closest to the HMMs using broad class models, as we only implement one detector per broad class. For HMMs using phoneme models, the resulting individual phoneme decoding is collapsed into a broad class transcription. Minor variations among the individual phones within each broad class may be captured in these otherwise redundant models, and the performance gain is significant.

Minimum string edit distance alignment was performed for all five systems. Table VIII summarizes the broad class transcription performance on all 1344 sx/i test sentences.[45] Our system accuracy exceeds that of both CI and CD HMMs using broad class models, which are of most similar complexity. We also exceed the accuracy of the context-independent HMM using phone models, though our system is using contextual constraints. Our system accuracy falls two points short of the context-dependent HMM using phone models. However, as we will discuss in the next section, there is vast room for improvement of our approach.

We find that the context-dependent HMMs attain significantly higher correctness rates relative to their context-independent counterparts. This, however, is at the expense of a significantly higher insertion rate, predominantly a result of not applying a language model for cleanup. Our landmark-

TABLE VIII. Broad class transcription performance for our landmark-based system vs various HMM approaches. HMMs result in high insertion rates in the pure speech recognition domain.

System	Accuracy	% correct	% ins	% del	% repl
Our System	70.3	76.0	5.7	11.3	12.7
HMM, CI/BC	65.5	68.4	2.9	17.6	13.9
HMM, CD/BC	65.1	90.5	25.4	1.4	8.1
HMM, CI/Ph	69.0	79.7	10.7	6.0	14.2
HMM, CD/Ph	72.2	91.5	19.3	1.6	6.9

based system is a conservative guesser, resulting in a low insertion rate similar to the context-independent HMMs. This is largely a result of landmark thresholding before decoding. Theoretically, the probabilistic decoding requires no thresholding, as low strength landmarks will have a correspondingly low $P(f|h=1)$. Decreasing or even removing the operating thresholds do increase correctness rates, but at the expense of insertions. We find that the accuracy levels are slightly higher after detector thresholding.

V. DIRECTIONS FOR FUTURE RESEARCH

In the preceding sections we have presented a landmark-based speech recognition frame-work fashioned on the principles outlined in Sec. I. Our broad class recognizer based on these ideas is competitive with equivalent HMM-based systems. Still, our implementation involves several design choices that are not necessarily optimal, leaving significant room for improvement of our computational approach. The immediate focus of our future research is to translate the ideas contained in this paper to practical speech recognition tasks, of which full phonetic recognition (discussed in detail in the following) is one example. However, history demonstrates that for all successful new approaches to the recognition problem, state-of-the-art performance was eventually attained by incremental advances in the various modules of the initial framework. Thus, in the following, we also discuss in some detail several areas for improvement, along with possible solutions, each addressing a limitation of our current implementation.

A. Full phonetic recognition

There are two distinct strategies to accomplish full phonetic recognition using the ideas developed in this paper.

1. Modeling phonetic sequences

This paper outlines a possible approach to map firing patterns of detectors into sequences over a symbolic inventory. Therefore, there are two classes of objects that are worth distinguishing. The first is $\mathcal{D}=\{d_f(t)|f\in\mathcal{F}\}$, a collection of feature detections. Here $\mathcal{F}$ is a family of feature types and for each feature type, $d_f(t)$ is a detector firing pattern associated with that type. The second object is a finite symbol set Σ and sequences over this symbol set, i.e., elements of Σ^*. In the system we have implemented, we chose $\Sigma=\{\mathrm{V},\mathrm{N},\mathrm{A},\mathrm{F},\mathrm{P},\mathrm{sil}\}$ (i.e., the set of broad classes), while $\mathcal{F}$ corresponds to the small number of leaf nodes in the distinctive feature hierarchy of Fig. 2.

In order to transition to full phonetic recognition, we need to let Σ be the set of phonetic units and, correspondingly, we will need to consider a more exhaustive set $\mathcal{F}$ of features. As long as the detectors for the set $\mathcal{F}$ and their firing patterns have sufficient discriminative power to separate all the phonemic classes from each other, our approach is in principle applicable to full phonetic recognition. There are two challenges. The first is to find an adequate set of feature detectors. The second is to deal with the possible combinatorial explosion one might anticipate if one were searching over all phonetic sequences rather than broad sequences alone.

Fortunately, neither challenge is insurmountable. In particular, our recent work[46] demonstrates the feasibility of scaling up to full phonetic recognition. In the implementations we have experimented with, we chose $\mathcal{F}$ to be a set of phonetic units and trained the corresponding phone detectors using methods from statistical learning. This larger set is sufficient to achieve phonetic recognition accuracy rates competitive with basic HMM systems.

Regarding the combinatorial challenge, it turns out that given the syllable-sized analysis units proposed in this paper, one only has to search over highly constrained phone sequences. Note that an important part of our strategy is to chunk the signal into syllabic nuclei, intervocalic sonorant sequences, and intervocalic obstruent sequences. Thus, for broad class recognition, we only need to consider possible broad class sonorant sequences and broad class obstruent sequences that can possibly lie between two adjacent vowels. These numbers are very small (12 and 42 in TIMIT, respectively, including silences). Our phone recognition experiments on TIMIT have shown that when one considers intervocalic sonorant and obstruent *phone-level* sequences, the numbers that occur are surprisingly limited (61 and 385, respectively, not including silences).

Thus, while there are two broad sonorant consonantal classes (nasal and approximant) that may combine to $32=2^5$ possible strings of length less than or equal to 4, only 12 actually occur in TIMIT. Similarly, though there are ten sonorant consonants and therefore 10^5 possible strings at the phonetic level, only 61 actually occur. These numbers illustrate that phonotactic and phonological constraints of the language dramatically reduce the set of possible consonantal sequences one needs to consider given the high-level chunking into syllable-sized units. Note that the statistics collected from TIMIT provide only an approximation to the distribution of phonological sequences in natural speech. However,

it seems evident that naturally occurring sequences will remain (with high probability) an exceedingly small subset of logically possible sequences.

In light of these experiments, it is clear that the principles outlined in this paper can be adapted to yield a viable phonetic recognition strategy. On the other hand, the approach explored in Ref. 46 is by no means the optimal execution of our high-level principles. For instance, a phone detector set could be replaced with a more general set of acoustic property detectors. Similarly, the search through possible sequences could be managed through various pruning or indexing strategies.

2. Transcription refinement

A second strategy is to perform transcription refinement of the broad class transcription provided by the system described in this paper. This involves expanding the distinctive feature hierarchy of Fig. 2 to include children of the current leaves that make distinctions between the individual phonemes within each broad class in a manner similar that suggested by Ref. 36. For example, adding place and voicing nodes under the stop leaf is adequate to distinguish between p [+labial, −voiced], b [+labial, +voiced], t [+alveolar, −voiced], d [+alveolar, +voiced], k [+velar,−voiced], and g [+velar,+voiced]. Determining additional feature values for transcription refinement can be accomplished using SVMs or any appropriate machine learning method.

In this new setting, the role of the current system is to provide landmarks around which these further features may be determined in a context-dependent way. Recall that the H variables of Sec. III C were introduced precisely for this reason. However, a complication arises from our probabilistic segment decoding method, which is capable of predicting a sequence for which there were degenerate detections (i.e., multiple correct candidates) or even no detections at all (this is rare). In these cases, it is not entirely clear where the true point for further analysis lies. When there are degenerate detections, an average of the landmark times weighted by their strength seems a reasonable choice. When a landmark is missing, we could take the maximal point of the prior time distribution, which amounts to the model's best estimate for the predicted context. Still, it may be the case that given a postulated broad class sequence with insufficient landmark information, more significant analysis must be performed. However, we believe improvements discussed earlier will not only improve broad class transcription, but also the selection of correct landmarks. Such an improvement will minimize this type of complication involved in transcription refinement.

B. Probabilistic sonority segmentation and vowel landmarks

Our method of probabilistic segment decoding (PSD) provides a means to accommodate error-making detectors to produce a list of likely transcription sequences for each obstruent or intervocalic segment. In contrast, the first two stages of our architecture, sonority segmentation and vowel landmark detection, are hard decisions. Therefore, mistakes made by these modules cannot be recovered. Furthermore, since they determine the obstruent and intervocalic regions for PSD, their errors propagate through later stages. This results in two significant performance bottlenecks.

The simplest approach to minimizing these bottlenecks is to improve the hard decisions they make. In the current scheme, this means improving the SVM performance, possibly through alternative signal representations, or even using alternative machine learning techniques. For example, neural networks have also been shown to be useful in this domain (see Ref. 47). Still, a more robust approach to solving the bottleneck problem is to transition to probabilistic sonority segmentations and vowel landmarks. Using such an approach, we could consider multiple candidate segmentations and vowel landmark sets for a given utterance.

This effectively amounts to performing the optimization of Eq. (1) over multiple candidate O^{seg} and O^{v}, in addition to B (see Sec. II D). Under this scheme, a less likely, but more accurate candidate segmentation can lead to better PSD performance and, ultimately, a more accurate transcription. A possible approach to implementing probabilistic vowel landmark detections arises from the dynamic baseline algorithm presented in Sec. III B 2. The number of landmarks chosen increases with the recursion depth d of the algorithm, resulting in one set V_d for each depth (i.e., $V_{d-1} \subset V_d$). The probability $P(V_d)$ of each set can be computed by

$$P(V_d) = \prod_{v \in V_d} P(h_v = 1 | f_v),$$

where $P(h_v=1|f_v)$ is the probability that the vowel landmark v is correct $(h_v=1)$ given its strength f_v. An approach for probabilistic sonority segmentation is not as immediate, though a scheme using variable thresholds may be useful.

C. Language model incorporation

Clearly, the above-presented broad recognition accuracies for HMMs are not the numbers we are used to seeing for state-of-the-art systems. In our experiments, we did not implement a phone- or word-level language model to clean up the phonetic transcription, resulting in lower broad class performance than is normally associated with HMM systems. The question remains of whether our framework would also admit comparable gains when a language model is incorporated.

We found in Secs. IV C and IV D that our probabilistic segment decoding method, while not always successful at choosing the correct sequence, provides exceedingly accurate N-best estimates. This means that if we impose higher-level linguistic constraints, we could have a superior chance at recovering the actual sequence from multiple top choices. Phoneme-, syllable-, or word-level n-gram models are the common choice for HMMs and would easily lend themselves to application within our probabilistic framework. Transitioning to probabilistic sonority segmentations and vowel landmark sets would result in even deeper language model benefits. However, it remains to be seen if the language model benefit for our system will exceed that of HMMs.

D. Landmark detector improvement

The individual landmark detectors are a major area for improvement that would lead to immediate overall performance gains. Reducing the number of detector insertions and deletions would put less burden on the integration procedure. Furthermore, more accurate landmarks would also sharpen the prior distributions, increasing the reliability of the posterior estimates.

One possible approach would be to implement acoustic parameters (APs) as an alternative to MFCCs. In our current implementation, the stop detector APs resulted in superior performance and computational efficiency. There is a significant body of existing research detailing the merits of APs over MFCCs for the broad classes (for examples, see Refs. 29, 31–33, 38, and 39). These parameters have been shown to increase robustness, exhibiting higher noise and speaker invariance.

Another approach is to individually address specific phoneme-level detector errors. One example is our stop detector's poor performance in [q] and [dx] detection. In this case, a possible remedy would be to create separate detectors for these problematic phones, resulting in multiple classifiers for each broad class. These classifiers could be integrated into a single broad class detector, using a logical OR of the subdetectors. This logic could be extended to a separate subdetector for every phoneme, resulting in complexity similar to HMMs using phoneme models.

There is also room for redesign of the detector set itself. In particular, we could implement broad class transition detectors, one for each ordered broad class pair. This would result in an augmented set of observables, but the probabilistic segment decoding formulation would remain exactly the same. A possible benefit of such an approach would be the sharpening of the prior time distributions. As it stands, a fricative detection, for example, can occur anywhere within the phoneme. Transition detectors, on the other hand, would be contained in much smaller regions, resulting in more pronounced distribution structure. However, it is unclear how the error rates of such detectors would fare to those currently in place.

The last approach for detector improvement would be to implement a different machine learning approach. SVMs have the nice property of providing a maximal separation between the two data classes, which helps reduce generalization error. However, the typically high number of support vectors involved in evaluation can be computationally taxing. Since the landmarks' detectors are independent modules of the system, we could implement any combination of machine learning methods here, so long as at the end of the day a series of landmarks are output. A study of various methods in this context will be required to determine suitable alternatives.

E. Alternative integration models

A final direction for system improvement is the probabilistic integration model. We have already touched on the possibility of alternative prior distribution estimation techniques. Another avenue is to explore the consequences of limiting the number of independence assumptions made in our probability model. More sophisticated detector time normalization methods may also be studied in order to sharpen the prior time distributions. The current approach of using the fraction of the obstruent or intervocalic region elapsed before a landmark is certainly better than using absolute times. However, for a stress-timed language such as English, there is still significant variation in the relative timing of phones in different contexts (see Ref. 48 for a discussion of the issues involved here).

The MAP approach developed in this paper is by no means the only statistical framework that can be employed (see Refs. 6, 49, and 50 for other examples). One possible alternative is to model the detector firings as Poisson processes with rate parameters dependent on the broad class sequence present and the region of that sequence you are in. For example, for the sequence "P F" the Poisson rate parameter for the stop detector would be high in the first half of the region and low in the second. While it is unclear if this approach would yield better results, it fits nicely with the point process representation provided by the detector hierarchy. It may also provide computational expression of neural coding theories of auditory processing.

[1]An example of this is our ability to learn new words, especially in a foreign language. In situations like this, we clearly have little or no access to morphological, syntactic, or higher level constraints, so our ability to recognize words in this case is an instantiation of pure speech recognition.

[2]J. Goldsmith, *Autosegmental Phonology* (Garland Press, New York, 1979).

[3]E. Sagey, "The representation of features and relations in non-linear phonology," Ph.D. thesis, MIT, Cambridge, MA, 1986.

[4]J. J. McCarthy, "Feature geometry and dependency: A review," Phonetica **43**, 84–108 (1988).

[5]K. H. Esser, C. J. Condon, N. Suga, and J. S. Kanwal, "Syntax processing by auditory cortical neurons in the FM-FM area of the mustached bat pteronotus parnellii," Proc. Natl. Acad. Sci. U.S.A. **94**, 14019–14024 (1997).

[6]Z. Chi, W. Wu, and Z. Haga, "Template-based spike pattern identification with linear convolution and dynamic time warping," J. Neurophysiol. **97**, 1221–1235 (2007).

[7]K. N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891 (2002).

[8]A. Juneja and C. Espy-Wilson, "Speech segmentation using probabilistic phonetic feature hierarchy and support vector machines," in *Proceedings of IJCNN*, International Joint Conference on Neural Networks, Portland, OR, July 20–24, 2003.

[9]M. Hasegawa-Johnson *et al.*, "Landmark-based speech recognition: Report of the 2004 Johns Hopkins Summer Workshop," in *Proceedings of ICASSP*, International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, March 18–23, 2005.

[10]J. R. Glass, "A probabilistic framework for segment-based speech recognition," Speech Commun. **17**, 137–152 (2003).

[11]S. Riis and A. Krogh, "Hidden neural networks: A framework for HMM/NN hybrids," in *Proceedings of the ICASSP*, International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, April 21–24, 1997.

[12]D. Huttenlocher and V. Zue, "A model of lexical access from partial phonetic information," in *Proceedings of the ICASSP*, International Conference on Acoustics, Speech and Signal Processing, San Diego, CA, March 1984.

[13]C. Demiroglu and D. V. Anderson, "Broad phoneme class recognition in noisy environments using the GEMS," in *Proceedings of the ACSSC*, Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 7–10, 2004.

[14]C. Espy-Wilson, T. Pruthi, A. Juneja, and O. Deshmukh, "Landmark-based approach to speech recognition: An alternative to HMMs," in *Proceedings of Interspeech*, Antwerp, Belgium, August 27–31, 2007.

[15]Not all 2^k possibilities may be instantiated as phonemes in a particular language. However, every phoneme in a natural language is one of these 2^k elements. In this sense, they constitute a universal inventory. Furthermore, the distinctive features allow one to characterize the set of natural classes and see why they are in fact a very small number of the logically possible subsets of 2^k phonemes. According to this theory, the natural classes are defined as all possible partial assignments to a k-dimensional boolean vector, admitting $O(3^k=\Sigma_{i=1}^{k}\binom{k}{i}2^i)$ possibilities. In contrast, without this boolean vector structure, one could potentially have $O(2^{2^k}=\Sigma_{i=1}^{2^k}\binom{2^k}{i})$ possible classes.
[16]M. Halle, "Phonology," in *Language*, edited by D. Osherson and H. Lasnik (MIT, Cambridge, MA, 1990), Vol. **1**, pp. 43–68.
[17]N. Chomsky and M. Halle, *The Sound Pattern of English* (Harper & Row, New York, 1968).
[18]M. Kenstowicz, *Phonology in Generative Grammar* (Blackwell, Oxford, 1994).
[19]G. A. Miller and P. Nicely, "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352 (1955).
[20]K. N. Stevens, *Acoustic Phoenetics* (MIT, Cambridge, MA, 1998).
[21]R. Jakobson, G. Fant, and M. Halle, *Preliminaries to Speech Analysis: The Distinctive Features* (MIT, Cambridge, MA, 1952).
[22]S. G. Parker, "Quantifying the sonority hierarchy," Ph.D. thesis, University of Massachusetts-Amherst, Amherst, MA, 2002.
[23]The sonority hierarchy arranges classes of sounds into equivalence classes arranged in decreasing order of sonority. This leads to the hierarchy Vowels > Approximants > Nasals > Fricatives > Affricates > Stops. Making a split at the Nasal/Fricative boundary realizes the binary distinctive feature [±son].
[24]One need not have a hard segmentation. One could postulate many candidate segmentations with a probability associated with each candidate segmentation. The associated probability may then be one of several considerations in settling upon an ultimate segmentation based on bottom-up and top-down analysis.
[25]The exceptions to this rule are syllabic nasals and liquids. These sounds are not protypical vowels in the acoustic sense, but function as syllabic nuclei nonetheless. In our implementation, we choose to regard them as sonorant consonants with the intent of improving vowel recognition performance at the expense of recognizing sonorant consonants, which compose a smaller fraction of the phonetic content of English.
[26]A vowel landmark is a point in the vowel at which its "vowelness" is most strongly exhibited. This corresponds to the peak of the sonority profile.
[27]S. Greenberg, H. Carvey, L. Hitchcock, and S. Chang, "Temporal properties of spontaneous speech—A syllable-centric perspective," J. Phonetics **31**, 465–485 (2003).
[28]Our use of the word classifier may lead to some confusion. Most machine learning methods produce a real-valued function that is thresholded to yield a binary-valued output. We use "classifier" to denote the real-valued function.
[29]P. Niyogi and M. M. Sondhi, "Detecting stop consonants in continuous speech," J. Acoust. Soc. Am. **111**, 1063–1076 (2002).
[30]C. Burges and P. Niyogi, "Detecting and interpreting acoustic features with support vector machines," Technical Report No. TR-2002-02, Computer Science Department, University of Chicago, Chicago, IL, 2002.
[31]C. Espy-Wilson, "Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in American English," J. Acoust. Soc. Am. **92**, 736–757 (1992).
[32]A. Salomon, C. Espy-Wilson, and O. Deshmukh, "Detection of speech landmarks: Use of temporal information," J. Acoust. Soc. Am. **115**, 1296–1305 (2004).
[33]T. Pruthi and C. Espy-Wilson, "Acoustic parameters for automatic detection of nasal manner," Speech Commun. **43**, 225–239 (2004).
[34]Y. Amit, A. Koloydenko, and P. Niyogi, "Robust acoustic object detection," J. Acoust. Soc. Am. **118**, 2634–2648 (2005).
[35]D. W. Tank and J. J. Hopfield, "Neural computation by concentrating information in time," Proc. Natl. Acad. Sci. U.S.A. **84**, 1896–1900 (1987).
[36]A. Juneja, "Speech recognition based on phonetic features and acoustic landmarks," Ph.D. thesis, University of Maryland, College Park, MD, 2004.
[37]This includes one energy and 12 cepstral coefficients, along with their delta and acceleration (double-delta) coefficients.
[38]O. Deshmukh, C. Espy-Wilson, and A. Juneja, "Acoustic-phonetic speech parameters for speaker-independent speech recognition," in *Proceedings of the ICASSP*, International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, May 13–17, 2002.
[39]Z. Xie and P. Niyogi, "Robust acoustic-based syllable detection," in *Proceedings of the ICSLP*, International Conference on Spoken Language Processing, Pittsburgh, PA, September 17–21, 2006.
[40]We will use Arpabet phonetic notation for the remainder of the paper to facilitate connections with experimental result figures.
[41]Note that the threshold for the vowel detector is not shown as it is not applied to the raw classifier output shown in the figure.
[42]In this sense, the use of the word "intervocalic" is technically a misnomer. However, to simplify discussion, we use it to refer to both segments bounded by two vowels and segments bounded by one or two sonorant–obstruent boundaries.
[43]T. Joachims, "Making large-scale SVM learning practical," in *Advances in Kernel Methods—Support Vector Learning*, edited by B. Schlkopf, C. Burges, and A. Smola (MIT, Cambridge, MA, 1999).
[44]K. Seymore *et al.*, "CMU Sphinx-3 English broadcast news transcription system," in *Proceedings of the DARPA Speech Recognition Workshop*, DARPA Broadcast News Transcription and Understanding Workshop, February, 1998.
[45]The reader might notice that the HMM broad class recognition accuracies quoted in Table VIII are in the ballpark of published phone recognition accuracies on the TIMIT database. However, such high phonetic accuracy quotes are the result of applying word-level language models or transition probability rescaling. To substantiate this claim, we reproduced a recent CI/Ph HMM experiment performed in Ref. 51 on TIMIT data that quotes a *phonetic* recognition accuracy of 67% when applying transition probability rescaling. Our implementation of their system led to similar phonetic recognition performance and translated into 81% *broad class* recognition accuracy, which is significantly higher than the numbers we quote for Sphinx-3. However, when the transition probability rescaling is removed from their decoding procedure, the broad class accuracy falls to 70%, which is consistent with our Sphinx-3 results.
[46]A. Jansen and P. Niyogi, "Point process models for event-based speech recognition," Technical Report No. TR-2008-04, Computer Science Department, University of Chicago, 2008.
[47]S. Chang, L. Shastri, and S. Greenberg, "Automatic phonetic transcription of spontaneous speech (American English)," in *Proceedings of the ICSLP*, International Conference on Spoken Language Processing, Beijing, China, October 16–20, 2000.
[48]E. Grabe and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," in *Laboratory Phonology*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin, 2003), Vol. **1**, pp. 515–546.
[49]W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown, "A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects," J. Neurophysiol. **93**, 1074–1089 (2005).
[50]N. Morgan, H. Bourlard, S. Greenberg, and H. Hermansky, "Stochastic perceptual auditory-event-based models for speech recognition," in *Proceedings of the of ICSLP*, International Conference on Spoken Language Processing, Yokohama, Japan, September 18–22, 1994.
[51]F. Sha and L. K. Saul, "Comparison of large margin training to other discriminative methods for phonetic recognition by hidden Markov models," in *Proceedings of the ICASSP*, International Conference on Acoustics, Speech, and Signal Processing, Honolulu, HI, April 15–20, 2007.

Developmental changes in the perception of pitch contour: Distinguishing up from down

Stephanie M. Stalinski, E. Glenn Schellenberg,[a] and Sandra E. Trehub
Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario L5L 1C6, Canada

(Received 6 February 2008; revised 3 June 2008; accepted 10 June 2008)

Musically untrained participants in five age groups (5-, 6-, 8-, and 11-year-olds, and adults) heard sequences of three 1 s piano tones in which the first and third tones were identical (A5, or 880 Hz) but the middle tone was displaced upward or downward in pitch. Their task was to identify whether the middle tone was higher or lower than the other two tones. In experiment 1, 5-year-olds successfully identified upward and downward shifts of 4, 2, 1, 0.5, and 0.3 semitones. In experiment 2, older children (6-, 8-, and 11-year-olds) and adults successfully identified the same shifts as well as a smaller shift (0.1 semitone). For all age groups, performance accuracy decreased as the size of the shift decreased. Performance improved from 5 to 8 years of age, reaching adult levels at 8 years. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956470]

PACS number(s): 43.75.Cd [DD] Pages: 1759–1763

I. INTRODUCTION

Subtle pitch changes play a critically important role in music. Melodies typically move in small steps, most commonly by one or two semitones (Vos and Troost, 1989). Moreover, major chords and harmonies, which signal positive emotions, differ by one semitone from minor chords and harmonies, which signal negative emotions (e.g., Gagnon and Peretz, 2003; Hunter *et al.*, 2008). Musically untrained listeners often notice pitch deviations that are smaller than one semitone, such as when a performer sings out of tune.

Distinguishing one pitch from another is less difficult than identifying one pitch as being higher or lower than another (Cooper, 1994; Johnsrude *et al.*, 2000; Sergeant and Boyle, 1980). The pitch contour of a melody—its pattern of successive changes in pitch direction (same, up, down)—is perceptually salient for listeners of all ages and levels of musical experience (for reviews see Dowling, 1994; Thompson and Schellenberg, 2006; Trehub, 2000). At times, infants fail to respond to transpositions of a relatively unfamiliar melody (i.e., change in pitch level, with contour and intervals preserved, Chang and Trehub, 1977; Trehub *et al.*, 1984), but they respond reliably to changes in its pitch contour (Trehub *et al.*, 1984), even when the directional changes are restricted to a single note (Trehub *et al.*, 1985). Although melodic contour is central to music perception regardless of age (e.g., Dowling, 1994; Trehub, 2000), little is known about the developmental course of contour processing. Frequency resolution is thought to be mature by about 8 years of age (Cooper, 1994; Maxon and Hochberg, 1982; Thompson *et al.*, 1999), but it is unclear whether progress in identifying directional changes in pitch follows a similar timeline.

Young children seem to have particular difficulty understanding concepts such as *high* and *low* (or *up* and *down*) in relation to pitch. Andrews and Madeira (1977) found, for example, that 6- and 7-year-olds could learn to associate a pitch of 262 Hz with a large pig and a pitch of 523 Hz with a small pig, but they were unable to designate the pitches as high or low (or *higher* or *lower*). By contrast, Jeffrey (1958) found that 5-year-olds had difficulty linking contrasting pitches with left and right button presses, which implies that 6-year-olds' successful acquisition of relations between pitch level and animal size (Andrews and Madeira, 1977) may have been facilitated by real world experience. Costa-Giomi and Descombes (1996) contend that 6-year-olds' difficulty identifying pitch direction is attributable, in part, to the multiple meanings of the terms high and low in English. When they trained French-speaking 6-year-olds to label pitches with the terms *aigu* and *grave*, which are used exclusively for high and low pitch, or *haut* or *bas*, which have multiple meanings (corresponding to the English terms high and low), children trained with the single-meaning terms were more accurate in labeling pitches separated by two octaves than were children trained on the multiple-meaning terms.

Although 6-year-olds readily discriminate ascending from descending pitch patterns, and they can match the patterns instrumentally, they do not spontaneously use the terms up, down, high, or low when referring to the patterns (Hair, 1977). With extensive training over the course of six days, Soderquist and Moore (1970) found, however, that 5-, 7-, and 9-year-olds improved substantially in their ability to identify the pitch direction of two pure tones (sine waves) differing in pitch. The minimal pitch difference required for correctly judging pitch direction was smallest for 9-year-olds and largest for 5-year-olds, who also showed the greatest benefit from training. It is possible, however, that children would achieve comparable or greater success without such extensive training in the context of stimuli with greater ecological validity.

The goals of the present study were twofold. Because of contradictory findings regarding 5-year-olds' ability to identify pitch direction (Van Zee, 1976; Soderquist and Moore, 1970), it was of interest to examine this ability in the context of ecologically valid stimuli (piano tones) and an age-appropriate task. The second goal was to examine age-

[a]Author to whom correspondence should be addressed. Electronic mail: g.schellenberg@utoronto.ca

related changes in the ability to identify directional changes in pitch. As noted, identifying directional changes in pitch is more difficult than differentiating pitches even though the former skill depends critically on the latter skill. As the frequency difference between two tones becomes smaller, one would expect tone differentiation to become more effortful (i.e., involving more cognitive resources), with negative consequences for the more challenging pitch direction task. In other words, identifying pitch direction should get progressively more difficult as the frequency difference between tones decreases. Moreover, one would expect the identification of pitch direction to be more difficult for listeners with immature frequency resolution than for those with mature frequency resolution. Because 8-year-olds are thought to be adultlike in terms of frequency resolution (Cooper, 1994; Maxon and Hochberg, 1982; Thompson *et al.*, 1999), they were expected to perform as accurately as adults at identifying pitch direction. Younger children were expected to perform more poorly because of their poorer frequency resolution and their poorer understanding of relational concepts (Andrews and Madeira, 1977; Hair, 1977; Jeffrey, 1958).

On the basis of well-documented individual differences in pitch resolution, even in musical contexts (Drayna *et al.*, 2001), individual differences in proficiency were anticipated at all age levels. Finally, because musical training enhances the ability to identify directional changes in pitch (e.g., Sergeant, 1973), participation was limited to those without musical training.

II. EXPERIMENT 1

A. Method

1. Participants

Twenty-six 5-year-olds (13 boys, 13 girls) were recruited from the local community. None had taken music lessons, and none had a personal or family history of hearing problems. Each child received a token gift for their participation. An additional two children were tested but excluded from the sample because they did not complete the entire testing session.

2. Apparatus and stimuli

Testing was conducted in a double-walled sound-attenuating booth (Industrial Acoustics Co.). Stimulus presentation and responses were controlled by a Dell personal computer with software created in REALBASIC via a Harman/Kardon 3380 amplifier and Electro-Medical Instrument Co. loudspeakers. The stimuli were presented at a comfortable volume.

The stimuli were modeled after those used to study adults with impaired pitch and music perception (Hyde and Peretz, 2004). A set of 11 tones was synthesized using *The Grand* (a piano timbre) and CUBASE software. The set included one standard tone with a fundamental frequency of 880 Hz, as well as five higher and five lower tones displaced in pitch by 4, 2, 1, 0.5, and 0.3 semitones upward and downward from the standard. Individual tones were 1 s in duration with a natural piano onset (i.e., rapid and slightly percussive) and a 10 ms decay. They were combined using SOUNDEDIT software into sequences of three tones. In each sequence, the first and third tones were standards, and the second tone was higher or lower than the standard. Thus, the stimulus set comprised ten different three-tone sequences, each 3 s in duration. To eliminate potential amplitude cues to pitch level (Grau and Kemler-Nelson, 1988; Neuhoff *et al.*, 2002; Neuhoff *et al.*, 1999), the amplitude of the three tones in each sequence was varied subtly and at random. Specifically, one tone was normalized at 94% of maximum amplitude, another at 97%, and the third at 100%.

3. Procedure

Children were first familiarized with the testing booth and told that they would be playing a game. The game involved hearing three sounds in a row and deciding if the second sound went up or down. To facilitate understanding of the two-alternative forced-choice (2AFC) task, the experimenter explained the relevant distinction in a number of ways, providing graphic as well as vocal examples. Included in this explanation was an animation of a ball rising or falling in synchrony with rising or falling pitch (by four semitones) to provide a visual analog of the task. Participants completed six training trials that included the animated "bouncing" ball. The children then completed six practice trials with the four-semitone change but no visual aid.

The test trials were presented in five blocks, beginning with the largest pitch change (four semitones) on the first block and proceeding through successively more difficult levels to the smallest change (0.3 semitones). Twelve trials were completed in each block, with the direction of the pitch change (i.e., up or down) determined randomly on each trial, constrained so that middle tone was higher than the first and third tones on six trials but lower on the other six. A flashing star was provided as feedback for correct answers. Stickers were given to the children at the end of each block to maintain their interest in the task.

B. Results and discussion

Preliminary analyses revealed that gender did not influence performance or interact with the difficulty of the testing block, either in the present experiment or in experiment 2. Gender was not considered further. On the training trials with a visual cue accompanying the pitch changes, 5-year-olds were near perfect (95% correct) in distinguishing up from down ($M=5.69$ correct out of 6 trials, SD=0.62). Performance on the six practice trials was considerably more variable, averaging 75% correct ($M=4.50$, SD=1.48), although 20 of the 26 children had at least four correct. Children's performance on test trials was compared with chance levels (50%, 6 out of 12 correct) by means of one-sample *t*-tests for each of the five blocks. Performance exceeded chance levels for each block, $ps<0.01$, with the mean exceeding 60% correct in each case and 80% correct on the first (easiest) block. A repeated-measures analysis of variance (ANOVA) confirmed that performance varied across blocks, $F(4,100)=9.64$, $p<0.001$. As shown in Fig. 1, performance decreased monotonically across blocks as the task became more difficult.

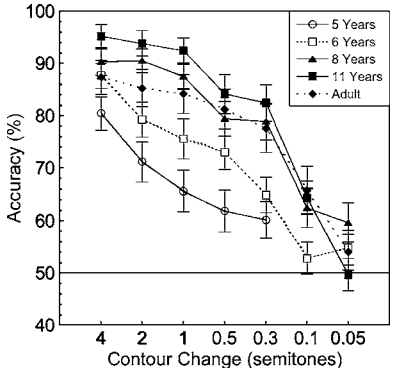


FIG. 1. Mean accuracy scores and standard errors for each age group and condition. In experiment 1, 5-year-olds were tested in five conditions. In experiment 2, older children (6-, 8-, and 11-year-olds) and adults were tested in seven conditions.

Also of interest was whether individual children performed at above-chance levels ($\alpha=0.05$, binomial test), corresponding to 9 or more correct responses out of 12 trials (75% correct). The results are summarized in Table I. Even with this conservative criterion, significantly more than half of the 5-year-olds (69%, or 18/26) succeeded in the four-semitone block, $\chi^2(1,N=26)=3.85$, $p<0.05$, which provides conclusive evidence that 5-year-olds can map the terms up and down onto changes in pitch. In each of the more difficult blocks, less than half of the children met this performance criterion.

Finally, because descending pitch changes are more frequent than ascending changes in actual music (Vos and Troost, 1989), the possibility of biased responding was investigated. One-sample *t*-tests were used to determine whether the number of up responses differed significantly from 6 (i.e., 50% up responses, 50% down responses), separately for each of the five blocks. There was no evidence of such a bias, with the proportion of up responses varying minimally from 50% across blocks (i.e., from 48% to 52%).

III. EXPERIMENT 2

A. Method

1. Participants

The sample included 29 6-year-olds (17 boys, 12 girls), 30 8-year-olds (13 boys, 17 girls), 30 11-year-olds (20 boys, 10 girls), and 29 young adults (10 men, 19 women). One additional 6-year-old was recruited but failed to complete the test session. No participant had formal music lessons or hearing problems. As in experiment 1, the children were recruited from the local community and received a token gift. The adults were university students who received either partial course credit or token remuneration.

2. Apparatus and stimuli

The apparatus and stimuli were identical to those from experiment 1 except that there were four additional tone sequences with more subtle changes in pitch to the middle tone. The middle tone of these additional sequences was displaced upward and downward from the standard tone by 0.1 and 0.05 semitones.

3. Procedure

For the children, the procedure was identical to experiment 1 except that (1) only the 6-year-olds played with stickers between blocks of trials, and (2) two additional, more difficult blocks of trials were added to the test session so that there were seven blocks of test trials rather than five. As in experiment 1, trials were presented in blocks, with successive blocks increasing in difficulty from the easiest block (4 semitones) to the most difficult block (0.05 semitones).

The procedure differed slightly for adults, who were told simply that they would be listening to a number of sequences, each containing three tones, and that their task was to decide on each trial whether the second tone was higher or lower in pitch than the first and third tones by selecting the appropriate option on a computer touch screen. To clarify the distinction between higher and lower, a number of three-tone vocal examples with higher and lower middle tones were provided. Feedback (correct or incorrect) was provided after each test trial.

B. Results and discussion

As with the 5-year-olds in experiment 1, children in the present experiment had no problem distinguishing up from down when pitch changes were accompanied by a visual cue. Performances were 98%, 99%, and 100% correct for the 6-, 8-, and 11-year-olds, respectively. Mean levels of performance on the practice trials with no visual cues exceeded 86% correct (better than five out of six correct) for each

TABLE I. Percentage of participants performing significantly better than chance (i.e., 9 or more correct out of 12) in each testing block. Five-year-olds (experiment 1) were tested only in the first five blocks. The seventh, most difficult block (experiment 2), is excluded because group performance did not exceed chance levels for any age group.

	Experiment 1	Experiment 2			
Block	5-year-olds	6-year-olds	8-year-olds	11-year-olds	Adults
4 ST	69.2	86.2	93.3	96.7	82.8
2 ST	46.2	69.0	90.0	96.7	82.8
1 ST	26.9	65.5	83.3	93.3	79.3
0.5 ST	38.5	55.2	70.0	76.7	72.4
0.3 ST	34.6	41.4	63.3	80.0	62.1
0.1 ST	⋯	13.8	33.3	33.3	51.7

group of children. For the test trials, separate one-sample *t*-tests for each of the four age groups and each of the seven blocks were used to compare performance with chance levels of responding (Bonferroni corrected for four tests for each block). Descriptive statistics are illustrated in Fig. 1. As expected from the results of experiment 1, older participants exceeded chance levels for the first five (easiest) blocks, ps <0.001, with mean levels of performance exceeding 65% correct for each age group in each block. Indeed, for the 8-year-olds, 11-year-olds, and adults, the means exceeded 78% correct for each block. For the more difficult, 0.1-semitone condition, the three older age groups (8-year-olds, 11-year-olds, and adults) exceeded chance levels, ps <0.0125, but the 6-year-olds did not. In the most difficult, 0.05-semitone condition, all four age groups performed at chance levels (i.e., <60% correct). In short, the seven testing conditions encompassed a range of difficulty that reached and surpassed the limits of participants' abilities.

Differences among age groups (four levels) and blocks (six levels) were analyzed with a mixed-design ANOVA. The final, most difficult block (0.05 semitones) was excluded from the analysis because of the floor effect observed across age groups. The analysis revealed main effects of age, $F(3,114)=4.79$, $p<0.005$, and block, $F(5,570)=80.62$, $p<0.001$, but no two-way interaction, $p>0.2$. As shown in Fig. 1, the age effect stemmed from poorer performance for the 6-year-olds compared to the other age groups. Indeed, when the 6-year-olds were excluded from the analysis, the effect of age disappeared, $p>0.3$. The main effect of block stemmed from monotonic decreases in performance across blocks, which were evident across age groups. The single exception was the slight increase in performance among 8-year-olds from the four-semitone to the two-semitone block, although performance was virtually at ceiling (>90% correct) in both blocks.

The present sample of 6-year-olds was compared to the 5-year-olds from experiment 1 with a mixed-design ANOVA with one between-subjects factor (age: two levels) and one within-subjects factor (block: five levels, the same blocks completed by both age groups). The results revealed a main effect of age, $F(1,53)=4.67$, $p<0.05$; 6-year-olds outperformed 5-year-olds. There was also a main effect of block, $F(4,212)=22.08$, $p<0.001$, with performance decreasing monotonically for both groups as the task became more difficult, but no interaction between age and block, $F<1$.

As in experiment 1, individual performance was evaluated against chance levels for each condition. The results, summarized in Table I, reveal monotonic decreases in the number of participants succeeding as the task increased in difficulty, the sole exception involving the transition from the 0.5-to 0.3-semitone block for 11-year-olds. In other words, group decrements in performance were paralleled by decreasing numbers of individuals who met the performance criterion. Even when the majority of participants succeeded at the task, others performed relatively poorly, highlighting marked individual differences in a basic aspect of pitch perception.

Finally, the possibility of a response bias was examined by comparing the number of up responses to unbiased responding (50% up responses), separately for each age group and trial block (Bonferroni corrected for four tests for each block). As with the 5-year-olds in experiment 1, there was no evidence that listeners of any age were biased to respond up or down.

IV. DISCUSSION

The results indicate that 5-year-olds can identify directional changes in pitch, which is consistent with the findings of Soderquist and Moore (1970) but inconsistent with those of others (Jeffrey, 1958; Van Zee, 1976). The present findings differ from those of Soderquist and Moore (1970) in revealing successful pitch-direction identification after mere minutes rather than hours of training. Five-year-olds' lesser success in the previous study may be attributed to the use of pure tones rather than piano tones, the use of two-tone rather than three-tone patterns, and the absence of feedback during testing. The importance of feedback can be seen in the improvement that was evident from the practice trials (mean of 75% correct) to the first block of test trials (mean of 80% correct), which featured the same stimuli. Feedback is particularly important for 5-year-olds, who often have difficulty generalizing from training stimuli to test stimuli. For example, 5-year-old children who are trained to respond to piano tones separated by three octaves as high and low fail to transfer that training to piano tones separated by seven semitones (Jeffrey, 1958).

As can be seen in Fig. 1, all age groups performed above chance levels in the easiest condition (a 4-semitone change) and below chance levels in the most difficult condition (a 0.05-semitone change). As with other auditory abilities (e.g., frequency resolution: Allen *et al.*, 1989; auditory thresholds: Leninhan *et al.*, 1971; pitch discrimination: Cooper, 1994; Maxon and Hochberg, 1982; Thompson *et al.*, 1999), gradual, age-related improvement was apparent. Specifically, 5-year-olds performed more poorly than all other age groups, and 6-year-olds performed more poorly than older children. Nevertheless, 5- and 6-year-olds children identified the direction of pitch changes for differences as small as three-tenths of a semitone. By 8 years of age, children succeeded in identifying changes of one-tenth of a semitone.

What accounts for the precocious abilities observed in the current study in contrast to children's reported difficulty mapping the terms high and low, higher and lower, or up and down onto the dimension of pitch (Andrews and Madeira, 1977; Flowers and Costa-Giomi, 1991; Hair, 1977; Van Zee, 1976)? In addition to the use of ecologically valid stimuli and feedback, as noted above, instructions in the present study were simpler than those in some previous studies. For example, Hair (1977, p. 200) required first graders to determine whether "the sounds in the first group" (an ascending or descending sequence of five resonator bells) "move in the same way" as "the sounds in the second group." The children were instructed, further, to "mark the green word 'yes' on your paper" if the sounds move in the same way and to "mark the red word 'no' on your paper" if the sounds do not move in the same way. "Otherwise, mark the question mark." Even if first graders could comprehend the concepts up and

down in relation to pitch, they would likely be confused by these instructions, especially in the absence of feedback. Interestingly, children who failed to identify the correct pitch direction succeeded in duplicating the ascending and descending sequences on the resonator bells.

As predicted, 8-year-olds' performance did not differ from that of adults, which is consistent with suggestions of mature pitch resolution by 8 years of age (Cooper, 1994; Maxon and Hochberg, 1982; Thompson *et al.*, 1999). Presumably, younger children's difficulty with pitch differentiation added to their cognitive load, exacerbating their difficulties on the pitch-direction task. The random variations in amplitude, however subtle, may have increased task difficulty for 5- and 6-year-olds, whose performance is often affected adversely by variations in an irrelevant stimulus dimension (Bialystok and Niccols, 1989). In fact, a number of researchers have reported that young children erroneously use terms associated with loudness when describing the pitch differences between paired tones (Andrews and Diehl, 1970; Hair, 1981; Van Zee, 1976). Finally, it is possible that the terms high and low, with their multiple meanings, increased task difficulty for the 5- and 6-year-olds (Costa-Giomi and Descombes, 1996). Regardless of the perceptual and cognitive factors that depressed performance in 5- and 6-year-olds, the fact remains that they succeeded in identifying the direction of pitch change at better than chance levels, even for pitch changes smaller than a semitone.

ACKNOWLEDGMENTS

Funding provided by the National Sciences and Engineering Research Council of Canada. Jessica Chan, Darcelle Edwin, and Jane Campbell assisted with participant recruitment and testing.

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

Andrews, F. M., and Diehl, N. C. (**1970**). "Development of a technique for identifying elementary school children's musical concepts," J. Res. Music Ed. **18**, 214–222.

Andrews, M. L., and Madeira, S. S. (**1977**). "The assessment of pitch discrimination ability in young children," J. Speech Hear Disord. **42**, 279–286.

Bialystok, E., and Niccols, A. (**1989**). "Children's control over attention to phonological and semantic properties of words," J. Psycholinguist. Res. **18**, 369–387.

Chang, H. W., and Trehub, S. E. (**1977**). "Auditory processing of relational information by young infants," J. Exp. Child Psychol. **24**, 324–331.

Cooper, N. (**1994**). "An exploratory study in the measurement of children's pitch discrimination ability," Psych. Music. **22**, 56–62.

Costa-Giomi, E., and Descombes, V. (**1996**). "Pitch labels with single and multiple meanings: A study with French-speaking children," J. Res. Music Ed. **44**, 204–214.

Dowling, W. J. (**1994**). "Melodic contour in hearing and remembering melodies," in *Musical Perceptions*, edited by R. Aiello and J. Sloboda (Oxford University Press, New York, NY), pp. 173–190.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (**2000**). "Genetic correlates of musical pitch recognition in humans," Science **291**, 1969–1972.

Flowers, P. J. and Costa-Giomi, E. (**1991**). "Verbal and nonverbal identification of pitch changes in a familiar song by English and Spanish speaking preschool children," Bull. Coun. Res. Music Ed. **101**, 1–12.

Gagnon, L., and Peretz, I. (**2003**). "Mode and tempo relative contributions to "happy-sad" judgments in equitone melodies," Cognit. Emot. **17**, 25–40.

Grau, J. A., and Kemler-Nelson, D. G. (**1988**). "The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness," J. Exp. Psychol. Gen. **117**, 347–370.

Hair, H. I., (**1977**). "Discrimination of tonal direction on verbal and nonverbal tasks by first grade children," J. Res. Music Ed. **25**, 197–210.

Hair, H. I. (**1981**). "Verbal identification of music concepts," J. Res. Music Ed. **29**, 11–21.

Hunter, P. G., Schellenberg, E. G., and Schimmack, U. (**2008**). "Mixed affective responding to music with conflicting cues," Cognit. Emot. **22**, 327–352.

Hyde, K. L., and Peretz, I. (**2004**). "Brains that are out of tune but in time," Psychol. Sci. **15**, 356–360.

Jeffrey, E. (**1958**). "Variables in early discrimination learning: II Mode of response and stimulus difference in the discrimination of tonal frequencies," Child Dev. **29**, 620–623.

Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (**2000**). "Functional specificity in the right human auditory cortex for perceiving pitch direction," Brain **123**, 155–163.

Leninhan, J. M. A., Christie, J. F., Russel, T. S., Orr, N. M., Hamilton, M. D., and Knoz, E. C. (**1971**). "The threshold of hearing in school children," J. Laryngol. Otol. **85**, 365–385.

Maxon, A. B., and Hochberg, I. (**1982**). "Development of psychoacoustic behaviour: Sensitivity and discrimination," Ear Hear. **3**, 301–308.

Neuhoff, J. G., Kramer, G., and Wayand, J. (**2002**). "Pitch and loudness interact in auditory displays: Can the data get lost in the map?," J. Exp. Psychol., Appl. **8**, 17–25.

Neuhoff, J. G., McBeath, M. K., and Wanzie, W. C. (**1999**). "Dynamic frequency change influences loudness perception: A central, analytic process," J. Exp. Psychol. Hum. Percept. Perform. **25**, 1050–1059.

Sergeant, D. (**1973**). "Measurement of pitch discrimination," J. Res. Music Ed. **21**, 3–19.

Sergeant, D., and Boyle, J. D. (**1980**). "Contextual influences on pitch judgement," Psych. Music. **8**, 3–15.

Soderquist, D. R., and Moore, M. J. (**1970**). "Effect of training on frequency discrimination in primary school children," J. Audit. Res. **10**, 185–192.

Thompson, N. C., Cranford, J. L., and Hoyer, E. (**1999**). "Brief-tone frequency discrimination by children," J. Speech Lang. Hear. Res. **42**, 1061–1068.

Thompson, W. F., and Schellenberg, E. G. (**2006**). "Cognitive bases of music listening," In *MENC Handbook of Musical Cognition and Development*, edited by R. Colwell (Oxford University Press, Oxford, UK), pp. 72–123.

Trehub, S. E. (**2000**). "Human processing predispositions and musical universals," in *The Origins of Music*, edited by N. Wallin, B. Merker, and S. Brown (MIT Press, Cambridge, MA), pp. 427–448.

Trehub, S. E., Bull, D., and Thorpe, L. A. (**1984**). "Infants' perception of melodies: The role of melodic contour," Child Dev. **55**, 821–830.

Trehub, S. E., Thorpe, L. A., and Morrongiello, B. A., (**1985**). "Infants' perception of melodies: Changes in a single tone," Infant Behav. Dev. **8**, 213–223.

Van Zee, N. (**1976**). "Responses of kindergarten children to musical stimuli and terminology," J. Res. Music Ed. **24**, 14–21.

Vos, P. G., and Troost, J. M. (**1989**). "Ascending and descending melodic intervals: Statistical findings and their perceptual relevance," Music Percept. **6**, 383–396.

Structural acoustics of good and bad violins

George Bissinger[a)]
Physics Department, East Carolina University, Greenville, North Carolina 27858, USA

(Received 22 January 2008; revised 11 June 2008; accepted 19 June 2008)

Modal-acoustic radiation measurements on 17 "bad-to-excellent" quality-rated violins—including three-dimensional modal analyses of *Titian* and *Willemotte* Stradivari and *Plowden* Guarneri del Gesu violins to investigate *extensional* as well as flexural motions—were examined for quality-related trends, generally by contrasting the properties of "excellent" and "bad" violins. All violins tested showed the same five "signature" modes below 600 Hz, with no obvious quality trends for mode frequencies or total damping. Bad–excellent comparisons of band-/modal-averaged damping (total, radiation and internal), mobility, radiativity, directivity, fraction-of-vibrational-energy radiated, effective critical frequency, and radiativity profiles up to 4 kHz generally showed no significant difference; the only "robust" quality differentiator was the ~280 Hz, Helmholtz-type A0 cavity mode radiativity where excellent violins were significantly higher. Radiation and total damping of two old Italian violins appeared slightly higher than those for bad violins below 2 kHz, partly due to lower effective critical frequency and partly because of slightly lower mass. Stradivari violins showed the highest and lowest directivity of all instruments tested. The *Titian* and *Plowden* top plate flexural/extensional mobility ratios appeared correlated with their directivity. Extensional motion in the "bridge island" between *f* holes peaked near 2.4 kHz, coinciding with the BH peak and a bridge/bridge-island impedance ratio minimum. [DOI: 10.1121/1.2956478]

PACS number(s): 43.75.De, 43.40.At, 43.40.Rj [NHF] Pages: 1764–1773

I. INTRODUCTION

A robust relationship between perceived violin quality and various mechanical–acoustical parameters amenable to scientific measurement has been an elusive scientific goal for almost two centuries. However, in the 1980s the advent of experimental modal analysis, near-field acoustical holography, zero-mass-loading excitation-response transducers, finite element and boundary element method computational techniques along with CT scan technology (to provide shape and density information)—all relying on the concurrent, equally rapid development of the computer—provided entirely new, and comprehensive ways to characterize the violin's dynamic and material properties. For the first time in the history of violin research it became possible to understand the surface vibratory behavior and motion of air in the *f* holes in great detail based on individual normal mode characterizations, or statistically via modal or band averages, accompanied by the potential to simulate these very vibrations.[1–11] Each violin mode could be characterized by mode shape, frequency, total damping (similarly for its major substructures, the top and back plate), and character (corpus bending modes, cavity, coupled cavity corpus, etc.), as well as by its acoustic radiation properties, e.g., radiation efficiency.

Such extensive and detailed information has led to some interesting simplifications, e.g., irrespective of quality, all traditional violins (and a complete violin octet[7]), properly constructed and set up, have only five "signature" corpus (top+ribs+back) normal modes in the open string region—albeit sometimes tailpiece or neck-fingerboard substructures can couple to these modes, splitting them. The violin's open string region (196–660 Hz for A=440 Hz) is crucial to the sound of the violin and also where the lowest plate modes are most important. Above ~700 Hz when the violinist holds/plays the violin the total damping increases so much that mode overlap suggests more statistical modal-/band-average analyses. Combining modal analysis with far-field acoustic radiativity measurements expands the descriptive structural acoustics parameters to radiation efficiency, radiation damping, internal damping, fraction-of-vibrational-energy radiated, and effective critical frequency.[11]

Add to this expanded range of parameters true three-dimensional (3D) modal analyses over a wide frequency range to examine a very important—but almost completely neglected—area of violin vibratory behavior, viz., in-plane extensional motion versus flexural out-of-plane motion. (To date only one investigation of these motions has been made, at a few fixed frequencies in the 400–600 Hz range.[12]) Although extensional motion does not lead directly to acoustic radiation, extensional motion can be transformed into flexural motion (and vice versa) at boundaries or discontinuities. The use of curved shells rather than flat plates for the violin's top and back implies significant extensional motion, hence the intractable analytical problem of determining relative contributions of flexural versus extensional motion for complicated shapes like the violin was approached experimentally in a straightforward way to provide a direct insight into the way the violin partitions its vibrational energy, which in turn is linked to its radiative properties. Of exceptional importance in this regard is the boundary-discontinuity concentration in the "bridge island" between *f* holes, where the

a)Electronic mail: bissingerg@ecu.edu.

soundpost, bass bar, and the *f* holes themselves are situated in the very region where string energy enters the corpus through the bridge feet.

This work, summarizing almost 10 years of wide-ranging vibration and radiation measurements on 17 violins and a complete violin octet, combines one-dimensional (1D) calibrated modal analyses of 12 quality-rated violins (including bridge, tailpiece, and neck-fingerboard substructures), new 3D scans of three old Italian and one modern violin with calibrated acoustical scans over a sphere in an anechoic chamber for all 17 violins, and a CT scan of each violin for density-shape material information. This 17-violin database (hereinafter labeled VIOCADEAS[13]) was mined for possible "robust" empirical parameter–quality relationship trends using a variety of approaches.

II. EXPERIMENT

Previous publications covered all relevant 1D experimental details for the comprehensive violin measurement-simulation program VIOCADEAS (Refs. 9 and 13, and references therein). All vibration measurements utilized zero-mass-loading laser scans of mobility $Y(\omega)$ (complex ratio of velocity/force) at >550 points over the ribs, top–back plates, bridge, neck-fingerboard, and tailpiece (the latter three substructures scanned from two orthogonal directions, creating two-dimensional scans) up to 4 kHz. The substructure (top, ribs, back) spatial-average, mean-square surface-normal mobilities $\langle Y^2 \rangle$ were used here to compute the substructure-area-weighted rms corpus mobility $\langle Y_{\text{corpus}} \rangle$. All 1D–3D mobility—and acoustic radiativity $R(\omega)$ (complex ratio of pressure/force)—measurements used zero-mass-loading, force-hammer impact excitation at the G-string corner of the bridge of violins suspended "free–free" from thin elastics (support fixture damping, $\leqslant 5\%$ of total damping, was neglected in all analyses).

Far-field radiativity scans at 266 points over an $r = 1.2$ m sphere in an anechoic chamber were made for all 17 violins. An over-the-sphere average of the mean-square radiativity $\langle R^2 \rangle$ was used to compute the rms radiativity $\langle R \rangle$; top and back hemisphere radiativities $\langle R_{\text{top}} \rangle$ and $\langle R_{\text{back}} \rangle$, respectively, (in-plane microphone points dropped), were used to compute a rough measure of directionality, the directivity $\langle D(\omega) \rangle = \langle R_{\text{top}}(\omega) \rangle / \langle R_{\text{back}}(\omega) \rangle$.

All violins were measured in playing condition with undamped strings at tension (A=440 Hz) without chin or shoulder rest. Nine of the twelve 1D-scan violins (and the Curtin violin) also had top and back plate mode frequency information provided by the maker. Two of the violins had bent, not carved, plates. Finally, all violins were played by the same violinist (Ara Gregorian) for quality evaluation purposes, although time constraints in the 3D-scan experiment necessitated a more general evaluation procedure for those violins.

The entirely new 3D mobility scans reported here significantly broadened the scope of VIOCADEAS, examining *extensional* as well as flexural surface motion over a broad frequency range. Over the available $2\frac{1}{2}$ day measurement period the *Titian* Stradivari (1715) and *Plowden* Guarneri del Gesu (1735) had essentially complete corpus scans, plus partial scans on the *Willemotte* Stradivari (1734, back plate only) and Joseph Curtin (2006, top plate only), plus a few high-density-point scans at specific frequencies. Corpus 3D scans required much more time than the $3\frac{1}{2}$ h needed for complete 1D grid scans (top, back, ribs, plus neck-fingerboard, tailpiece, and bridge from two orthogonal directions), plus acoustic scans over a sphere because of the prior 3D surface geometry scans needed for each violin to accurately specify *XYZ* coordinates for all points in the 3D measurement grid. Three separate lasers simultaneously measured the surface velocity vector along each laser's beam direction at each point; top plate exclusion zones were somewhat larger than for 1D measurements due to neck-fingerboard, bridge, or tailpiece surface shadowing. The time-limited 3D experiment led to an error of omission in the automated backhemisphere radiativity scans, which covered 0–4 kHz, not 0–5 kHz as in all other 3D measurements. In this article, 1D–3D force hammer excitation was along the *X*-direction only.

The three mobility vectors were then decomposed into orthogonal components in a chosen frame of reference.[14] Since the violin has no flat surfaces, the *Y* direction (perpendicular to the "plane" of the violin)—the component used for comparison with previous 1D measurements on top and back plates—was labeled out-of-plane (OP), and the *XZ* plane labeled in-plane (IP) for convenience. In the bridge "island" between *f* holes, *X* and *Z* mobilities were analyzed separately to understand extensional motion in the crucial region where string energy enters the violin.

III. RESULTS

Mobility spectra provided normal mode frequency, total damping, ζ_{tot}, and mode shapes to characterize each violin's vibrations, whereas radiativity spectra provided directivity as well as radiativity "profiles" useful in characterizing violin sound. Radiation efficiency, R_{eff}, radiation damping, ζ_{rad}, *effective* critical frequency, f_{crit}, and the fraction-of-vibrational-energy-radiated $F_{\text{RAD}} = \zeta_{\text{rad}} / \zeta_{\text{tot}}$ were computed from combined mobility and radiativity measurements. Various approaches were used to examine experimental parameter–quality relationships, e.g., plots of parameter versus quality rating to look for trends, or clumping the violins into "bad," "good," and "excellent" groups for statistical analysis, or applying trend-line analysis to quality-grouped parameters, etc. At low frequencies where individual signature modes were seen for all violins irrespective of quality, mode frequencies and total damping were examined for trends. At higher frequencies mode overlap becomes so pervasive and mode shapes so variable, even for our free–free support, that more statistical band- or modal-average analyses and trend lines—notably to estimate f_{crit}—were utilized. Note that string-peak structures were still obvious and much narrower than the corpus peaks over the entire frequency range. The 3D vibration OP–IP measurements were examined for a possible link to radiation directivity $\langle D \rangle$.

Violin subjective quality ratings were on a 1–10, three-main-class rating scale—bad (1–3), good (4–7), and excel-

lent (8–10). Previously violin parameter versus quality comparisons were between the highest ranking good and the bad violins;[11] the new data on three old Italian violins has extended the quality range to excellent. Thus comparisons have evolved to excellent versus bad comparisons (plus all-violin averages where appropriate) in an attempt to magnify possible differences in quality-related parameters; note that *quality–class* comparisons are stressed rather than individual violin results. Because any analysis of only 17 violins suffers from small-number statistics for the various quality groups, only robust quality quantifiers—where standard deviation error bars did not overlap—will be given much consideration. Even robust quantifiers, however, cannot be considered reliable without extensive corroboration.

A. Mobility and radiativity

1. Signature modes

Our discussion of individual modes will be limited to just those five low-lying modes observed in all measured violins and the violin octet. These signature modes fall into two major classes:

(1) Cavity modes: (a) A0, always the lowest frequency mode, $f_{A0} \approx 280$ Hz. A Helmholtz-type mode characterized as a mass-plug oscillating under the influence of the cavity "spring," always a strong radiator, and (b) A1, the first longitudinal mode, sometimes an important radiator with $f_{A1} \approx 1.7 f_{A0}$. These are *coupled* modes and each has an admixture of the other, a circumstance that strongly affects the A0 volume dependence and the upper–lower bout pressure ratios.[15]
(2) Corpus modes: (a) CBR, the lowest frequency corpus mode with shearlike IP relative motion between top and back plates, a ‡ OP nodal line pattern on top and back plates accompanied by out-of-phase *f*-hole volume flows and thus relatively weak radiation, and (b) the first corpus bending modes $B1^-$ and $B1^+$—which both radiate strongly and also strongly through the *f* holes.[10]

The strongly radiating A0, $B1^-$, and $B1^+$ modes in the open-string, 196–660 Hz region are crucial to violin sound.

Figure 1 shows the rms corpus OP mobility $\langle Y_{corpus} \rangle$ and rms top hemisphere radiativity $\langle R_{top} \rangle$ for the *Titian* Stradivari and *Plowden* Guarneri del Gesu in the open string region with the signature modes annotated. These radiativity and mobility curves were not exceptional in magnitudes, widths, or peak placements compared to other violins.

2. Magnitudes

The mobility and radiativity magnitudes of bad, excellent, and 14- or 17-violin averages (nominally good violins) are presented in Fig. 2 as 250 Hz band averages, with two important exceptions—"A0" is an average over ±10 Hz around the A0 peak, while the 400 Hz band is averaged from 300 to 499 Hz to exclude A0, but including CBR and $B1^-$. All higher bands are over 250 Hz intervals; the band centered at 625 Hz always includes $B1^+$. This band choice ensures the lowest three radiativity bands are dominated by A0, $B1^-$, and $B1^+$, respectively, *Intraband* variations are shown

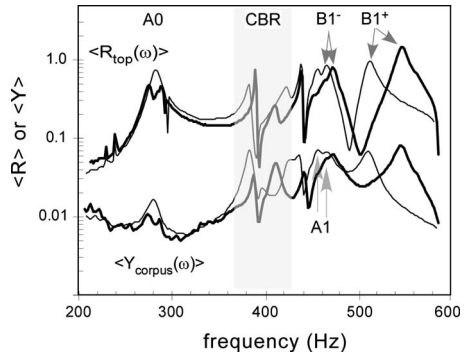


FIG. 1. Signature mode region OP (log) corpus mobility $\langle Y_{corpus} \rangle$ (lower curves, m/s/N) and top hemisphere (log) radiativity $\langle R_{top} \rangle$ (upper curves, Pa/N) for *Titian* Stradivari (thick line) and *Plowden* Guarneri del Gesu (thin line) vs frequency. Note low A0 mobility; narrow structures are string harmonics (always narrower than corpus peaks). *Titian* A0 shows neck-fingerboard or tailpiece coupling.

in Fig. 2 with standard deviation (s.d.) error bars. Radiativity provides an objective measure of how effectively forces applied at the bridge can be turned into sound without any ear sensitivity weighting. Figure 2 clearly shows that the only robust difference between bad and excellent (old Italian) violin radiativity occurs for A0. This does not imply that we may not *perceive* them as louder, or that their directivity is the same. In fact, both aspects are important.

The average mobility falls off smoothly above the maximum near 2.4 kHz, which was originally attributed to the bridge "rocking" about the waist, with the feet relatively fixed. A similar peak has shown up in every part of the energy chain: bridge driving point, averaged-over-bridge, bridge feet, averaged corpus mobility, and radiativity.[16] However experiments by Jansson and co-workers[17] with solid and standard bridges clearly demonstrated that this

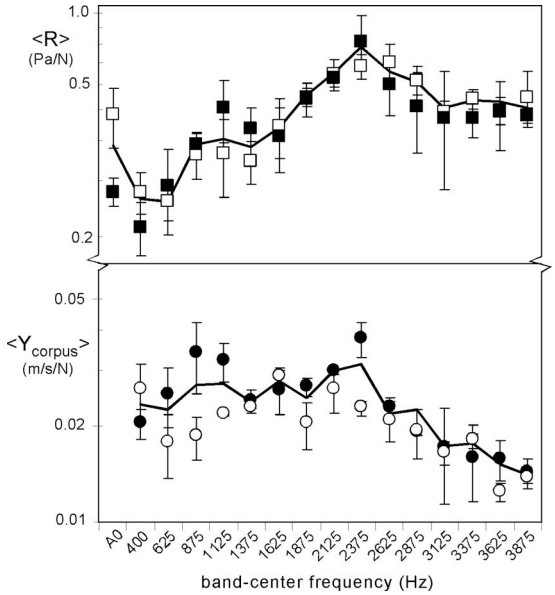


FIG. 2. Top panel: Band-average (log) radiativity $\langle R \rangle$ (17 violins); bottom panel: Band-average corpus OP (log) mobility $\langle Y_{corpus} \rangle$ (14 violins, A0 mobility omitted) vs band-center frequency: "bad"—solid symbol), "excellent"—open symbol (all old Italian), and average—line (here average is nominally "good"). (s.d. error bars reflect *intraband* variations only.)

peak did not originate in the bridge rocking about the waist, a conclusion corroborated by an experiment where frequencies for this type rocking were varied widely.[16] This matter of the BH peak will be addressed more thoroughly in a later section on 3D measurements. Corpus mobility showed some evidence of a bad–excellent difference in the 875–1125 and 2375 Hz bands.

The radiativity profile of the bad violins was somewhat more peaked than the excellent, with the A0 and high frequency ends both lower. Overall, excellent violins had a somewhat more uniform response.

B. Violin mode properties versus quality

The following results summarize violin normal mode properties sorted by their subjective quality rating. All violins were played by the same excellent violinist Ara Gregorian; the 12 VIOCADEAS violins' overall ratings used a systematic multiparameter rating scheme, while the Curtin, Zygmuntowicz, Stradivari, and Guarneri del Gesu violins in the 3D experiment were all evaluated in a different way and have numerical ratings supplied by the author based on Gregorian's comments while playing, listeners comments, and in case of the old Italian violins their historical summary status. This qualitative rating should not be considered absolute—in the sense of some other excellent violinist coming to exactly the same numerical value—but as a reasonably reliable evaluation based on a consistent rating scheme. The fact that no robust quality trends emerged from this analysis reflects a reality in trying to quantify violin quality.

1. Signature mode frequency and damping

Schleske, a prominent German violin maker who has been the leader in incorporating modal analyses into violin making, stated[5] that the frequency of $B1^{(+)}$ acts as a "tonal barometer" for violin sound, with frequencies <510 Hz leading to a "somewhat soft" violin with dark sound, lacking "resistance" to bowing. On the other hand frequencies >550 Hz were characteristic of "…'stubborn' violins with bright sound, possibly with a tendency to harshness, and with strong 'resistance' to the player." It is unclear how a 10% change in corpus $B1^+$ frequency could cause such a change in perceived mechanical response since the string terminations are relatively insensitive to corpus vibrations except in the case of wolf-tones. Such "mechanical" characterizations may actually have significant *acoustic* components. Rohloff, in a 1964 experiment where filter-controlled violin sound reached the violinist only through headphones, found that a violin's resistance was linked, not to the bowing force needed to initiate tones as one might expect, but rather to *acoustic strength above* 4 kHz: "easy-speaking" violins had extended response above 4 kHz, "hard-speaking" violins had limited response.[18]

Schleske's remarks, however, do suggest that $B1^+$ frequency might in some way be a convenient gauge of quality, a conjecture readily tested by plotting $B1^+$ and all other signature mode frequencies for all 17 violins versus subjective quality rating in Fig. 3. Overall, A0 averaged 275 ± 9 Hz (s.d. errors: min. 253 Hz, max. 282 Hz), A1 averaged 469 ± 19 Hz (430–494 Hz), CBR averaged 407 ± 31 Hz (363–452 Hz), $B1^-$ averaged 475 ± 16 Hz (439–500 Hz), and $B1^+$ averaged 541 ± 22 Hz (511–591 Hz).

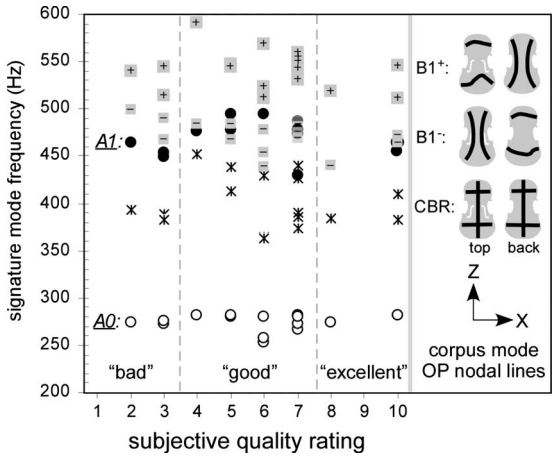


FIG. 3. Signature mode frequencies vs subjective quality rating for 17 violins: A0 (○), A1 (●); CBR (*), and $B1^-$ and $B1^+$ (shaded squares with − or +). Excellent violins all old Italian. Corpus mode OP nodal line patterns are on the right-hand side.

The A1–A0 frequency ratio was 1.71 ± 0.05, in close agreement with values obtained for a rigid violin-shaped cavity.[15] An important coupling between A0 and A1 discovered in the rigid-cavity experiment reduces the expected Helmholtz-type A0 volume dependence significantly, providing the physical basis for the difficulty encountered by Hutchins and Schelleng in reliably scaling the violin octet "main air" resonance frequency to the various different pitch ranges by adjusting rib heights.[19]

Figure 3 shows no obvious quality-related trends in signature mode frequencies. The two bent wood violins (6,7 ratings) were unexceptional. Considering the remarkable range of frequencies for the 6- to 7-rated good violins, the only reasonable conclusion to be drawn from Fig. 3 is that signature mode frequencies are not robust quality indicators. Similar scrutiny of signature mode total damping, extracted by peak-fitting routines, also showed no robust quality indicators.

2. Damping trends

In an earlier report the frequency dependence of the total damping was investigated by isolating three-violin subsets of good and bad violin normal mode damping values from the mobility spectrum fits.[11] Power-law trend lines of the form $\zeta_{\text{tot}} = C f^X$ were used to quantify damping falloff trends, with $x \approx -0.5$ seen for various structures.[20] Note that damping falloff for isolated top and back plates was consistent with $x \approx 0$, significantly slower than for the corpus.[9] Earlier analysis had shown a significant difference in x, with the good–bad exponent difference being 0.16 ± 0.07. Here the comparison, shown in Fig. 4, is between the *Plowden-Titian* and the bad violin data sets. The *Plowden–Titian* trend-line exponent, $x = -0.40$, was within error of the bad $x = -0.45 \pm 0.05$;[11] obviously total damping trends are *not* robust quality discriminators. Overall, total damping trends were similar for any violin quality class, with the earlier difference

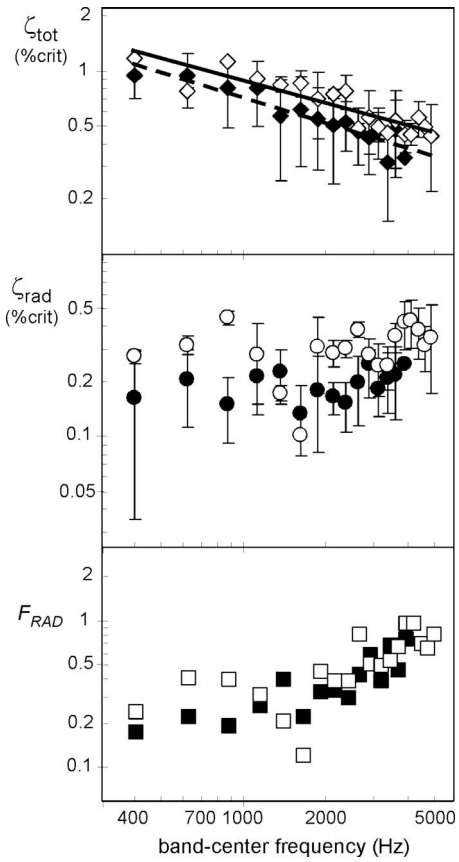


FIG. 4. Band-averaged (log) ζ_{tot} (top panel, with power law trend lines: Solid line "excellent," dashed line "bad"), (log) ζ_{rad} (middle panel), and (bottom panel) (log) F_{RAD} ($=\zeta_{rad}/\zeta_{tot}$) for excellent (open symbol) and bad (closed symbol) violin subsets vs (log) band-center frequency. *Intraband* s.d. error bars (top-middle); average F_{RAD} propagated errors (not shown) nominally ±40%.

most likely due to the paucity of data. Note that even though the *Plowden-Titian* total damping values were within error of the bad violin values for any band below 4 kHz (with one exception), they were always slightly larger, primarily due to a radiation damping difference.

The radiation damping was computed from R_{eff} ($\zeta_{rad} \propto R_{eff}/fM$, where M=violin mass, f=mode or band-center frequency). R_{eff} plateaus above f_{crit}, creating a "knee" in the ζ_{rad} frequency dependence at f_{crit},[11] a maximum in the fraction-of-vibrational-energy-radiated $F_{RAD}=\zeta_{rad}/\zeta_{tot}$, and the most efficient region for vibration–sound conversion.

Since $\zeta_{tot}=\zeta_{rad}+\zeta_{int}+\zeta_{fix}\approx\zeta_{rad}+\zeta_{int}$ (*if* ζ_{fix} can be neglected), knowing ζ_{tot} and ζ_{rad} provides the only reliable path to computing ζ_{int} for internal (heat) losses in a structure. (The support fixture damping ζ_{fix} was ⩽5% of ζ_{tot} for our free–free suspension and therefore neglected.) However when the violinist holds the violin, ζ_{fix} *dominates* ζ_{tot}. Figure 4 shows excellent (*Plowden-Titian*) radiation damping higher than bad (although still with overlapping error bars) except for the 1375–1625 Hz bands. Lower ζ_{rad} values observed for bad violins are consistent with higher critical frequencies *and* higher violin mass. Bad violin masses were ~10% higher than excellent.

Higher excellent ζ_{rad} did lead to higher F_{RAD} values across the frequency span even though ζ_{tot} was also larger; the difference was more apparent at the lower frequencies, essentially disappearing above 2 kHz, although propagated errors were so large (±40%) that little could be made of any difference. Interestingly, holding the violin makes support fixture damping (now the violinist!) dominate the total damping and leads to the violinist possibly perceiving a different situation when comparing excellent to bad violins; F_{RAD} decreases as expected (~50%), but the *relative difference* between violins increases.[9] Note that F_{RAD}—the "egress" filter for vibration–sound energy conversion—is completely independent of the "gatekeeper" filter, the violin bridge, which intermediates the initial vibrational string-corpus input energy transfer. It is an interesting "coincidence" that F_{RAD} peaks near those frequencies most strongly affected by bridge rocking mode frequency changes,[16] and where the ear is most sensitive.

The internal damping of the violin, $\zeta_{int}\approx\zeta_{tot}-\zeta_{rad}$, falls off with frequency somewhat faster than ζ_{tot} since ζ_{rad} increases slowly up to f_{crit}. Because internal damping is similar between these extreme violin quality classes—and propagated errors so large—no definite statement is possible. Practically speaking, at $f>3$ kHz, heat losses from air and surface absorption effects in a large auditorium—to say nothing of the expected violinist "support fixture" damping—are likely more important than internal damping.

3. R_{eff} and effective critical frequency

Due to its insensitivity to any shape-material properties of the vibrating object, R_{eff} becomes a very useful structural acoustics parameter to quantify vibration–radiation conversion. For a particular experimental setup, R_{eff} varies only with the ratio $\langle R^2\rangle/\langle Y^2\rangle$ for each mode, thus—guided by baffled piston radiation—implying an f^2 dependence and second-order polynomial trend line. Since violin shape and materials make accurate critical frequency estimates impossible, and the orthotropic nature of wood gives two values, *effective* critical frequencies f_{crit}, were estimated initially from experimental R_{eff} trend lines[11] solved for $R_{eff}=1$. Mode-to-mode R_{eff} varies widely, however, so that even in successive 250 Hz bands with two to four modes, substantial adjacent-band jumps were common, leading in turn to occasional unreliable second-order polynomial fits.

This difficulty was mostly circumvented by using band-averaged $\langle R\rangle/\langle Y\rangle$ plots to "linearize" the data; for the VIOCADEAS setup f_{crit} was determined by solving the trend-line equation for $\langle R\rangle/\langle Y\rangle=35.2$ Pa s/m. Normally linear and second-order polynomial trend-line f_{crit} values were averaged. For f_{crit} values differing by more than 5%, an exponential trend-line value was added and the average recomputed.

The band-average $\langle R\rangle/\langle Y\rangle$ data for bad and excellent violins in Fig. 5 shows a distinct difference across the entire range, with the excellent linear trend line crossing the 35.2 Pa s/m line at ~3.5 kHz, while the bad trend line crosses near ~4.3 kHz, defining f_{crit} for each quality class. The overall average f_{crit} violin value, 3.9 kHz, was in good agreement with 4.5–4.9 kHz values computed by Cremer for

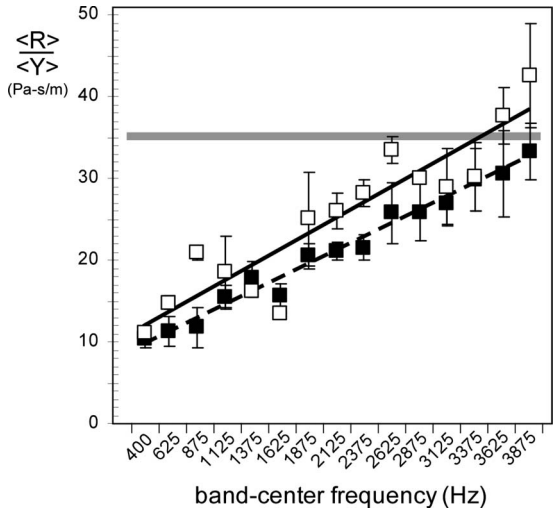


FIG. 5. Effective critical frequency estimates from 250 Hz band-averaged corpus $\langle R\rangle/\langle Y\rangle$ for two "excellent" (□) and three "bad" violins (■) vs band-center frequency, with linear trend lines (excellent, solid line, bad, dashed line). (s.d. error bars reflect *intraband* variations only). Lowest band center at 400 Hz. VIOCADEAS setup: f_{crit} ($R_{eff}=1$) at $\langle R\rangle/\langle Y\rangle = 35.2$ Pa s/m (broad gray line).

violin-size flat rectangular spruce and maple wood plates (cross-grain), with along-grain f_{crit} values two octaves higher.[21]

A dip near 1625 Hz, a universal aspect of violin R_{eff} curves, is suggestive of a link to the ring frequency for cylinders,[20] for violins nominally 1 kHz.[13] Relative prominence for this band is associated with "nasality" in the overall tone. Higher excellent violin ζ_{rad} in Fig. 4 follows directly from increased $\langle R\rangle/\langle Y\rangle$ and the resultant lower f_{crit}, in combination with lower mass.

A plot of f_{crit} versus quality rating for 14 violins is presented in Fig. 6. Looking only at bad versus excellent violins, a case might be made that there is a significant difference between these classes, but the best of the good violins show a range encompassing these quality extremes, again undermining any robust quality-related trend. Note also that machine-figured plates in factory violins typically run significantly thicker than these tested bad violins, implying a lowered critical frequency.

4. Directivity

Although Fig. 2 showed little difference in averaged-over-sphere radiativity between bad and excellent violins, the

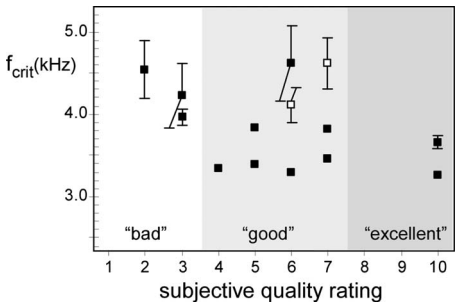


FIG. 6. f_{crit} vs subjective quality rating for 14 violins (□—bent wood violins, all others carved plates). (s.d. errors reflect variation between various trend-line estimates only; point size hides smallest errors.)

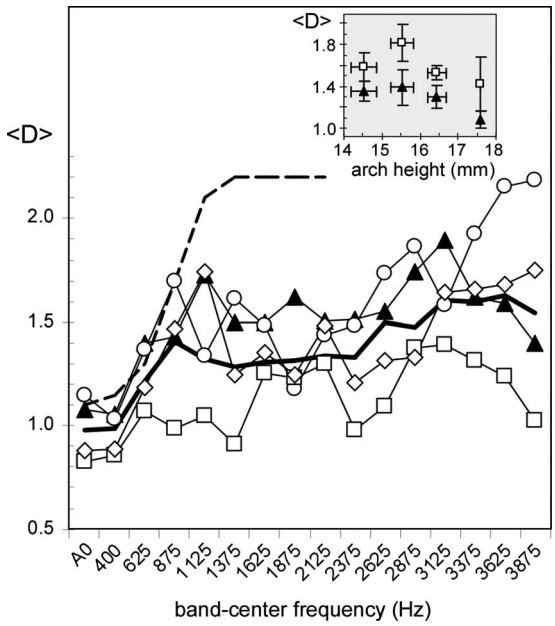


FIG. 7. Directivity for "bad" (▲), *Titian* (○), *Plowden* (◇), *Willemotte* (□), and 17-violin average (thick line) vs frequency. Directivity of *f*-hole radiation only (Ref. 10) shown as dashed line. *Titian* had highest $\langle D\rangle$, *Willemotte* had the lowest. Nominal intraband variations 15%. [Inset: Average $\langle D\rangle$ for 625–875 Hz (▲) and 3125–3875 Hz (□) regions vs arch height for 17 violins; s. d. error bars.]

way the violin is always held by a soloist implies that violin sound directionality must be an important facet of being heard over the orchestra. The violinist would be expected to hold the violin in the way that most effectively gets the sound to the audience, implying the top radiates more effectively than the back. The sound from the back also effectively has two extra (floor–back wall) bounces before heading into the hall, further diminishing its importance. The directivity $\langle D\rangle=\langle R_{top}\rangle/\langle R_{back}\rangle$ summarized for all 17 violins in Fig. 7 is a simple measure of sound directionality. Bad violin $\langle D\rangle$ generally was above the 17-violin average; the *Titian* had the highest $\langle D\rangle$ overall of any violin tested to date (with one bad violin quite close) and the *Willemotte* the lowest, while the *Plowden* was very close to average. Obviously directivity varies widely, even among the violins of one maker, with no link to perceived quality.

Figure 7 highlights an unexpected behavior, viz., the surprisingly fast rise of $\langle D\rangle$ at 625–875 Hz (followed by a plateau from 1 to 2.5 kHz, then a slow rise above 2.7 kHz). In the 625–875 Hz region where $\lambda \geqslant$ violin size, $\langle D\rangle \approx 1$ (isotropic) radiation would be expected. A reasonable interpretation of this rapid rise, based on recent "patch" NAH results for just *f*-hole radiation[10] compared to the anechoic chamber measurements of corpus+*f*-hole radiation, is based on the fact that *f*-hole-only radiation contributed ~50% to the overall violin radiation at $f<1$ kHz (falling off with increasing frequency) *and* was significantly more directional at lower frequencies (dashed line, Fig. 7) than expected, e.g., if the 625 or 875 Hz bands had a 50–50 *f*-hole-corpus radiation balance, then the amalgamated $\langle D\rangle$ would be $(1.3+1)/2 \approx 1.2$, or $(1.8+1)/2 \approx 1.4$, respectively, close to 17-violin values.

Does arching affect directivity? The *Willemotte* top plate arching of 17.6 mm was highest of all violins tested (along with the Curtin violin), while the *Plowden* was the next-to-lowest at 14.1 mm. CT scan bridge slices were used to estimate arch heights ±0.3 mm; Curtin and Zygmuntowicz violins had directly measured values. Arching groups of 14 to 15 mm (seven violins), 15 to 16 mm (three violins), 16 to 17 mm (five violins), and 17 to 18 mm (two violins) were plotted versus average $\langle D \rangle$ in the 625–875 and 3125–3875 Hz bands to create the inset in Fig. 7. Overall, the highest arch violins appear to have lower overall directivity, with the suggestion of a maximum in the 15 to 16 mm range. However, real arching effects might be entangled with OP–IP differences that also imply a directivity link.

C. Three-dimensional modal analyses

The 3D modal analyses examined a difficult and hence neglected area of violin vibratory behavior, viz. out-of-plane flexural versus in-plane extensional motion. The fundamental coupling between flexural and extensional motion, combined with possible transformations of extensional into flexural motion—and vice versa—at edges (e.g., rib joints) and discontinuities (e.g., bass bar, *f* holes, soundpost), make this an extraordinarily difficult analytic problem for complicated structures. However, experimentally determining the relative importance of flexural versus extensional motion is a promising avenue for understanding violin radiative properties at a fundamental level. These measurements investigated two important aspects of violin sound: (1) how OP–IP vibrational energy partitioning might affect violin radiation, and (2) how violin corpus motion immediately in the vicinity of the bridge feet might relate to string energy transfer through the bridge feet to the corpus.

1. OP–IP vibrations and directivity

OP mobility extracted from 3D scans was shown in Fig. 1 for the *Titian* and *Plowden*. The magnitudes were similar to those of other violins. Of more interest here is the strength of OP relative to IP motions. Since only the OP component is responsible for radiation, the reasoning here—assuming the same averaged overall magnitude—is that larger OP/IP ratios should correlate with more sound production. Figure 8 presents OP/IP ratios for the top plate and back plates of all violins with 3D scans, even partial ones. Three maple backs all had similar OP/IP ratios, implying a constant radiative contribution, while the top plate OP/IP—with varied-geometry bass bars and *f* holes—differed significantly. (Violin-making lore has long stressed the relative importance of the top versus back to violin sound.)

Maple backs had much higher OP/IP ratios than spruce tops, and the *Titian* top plate had a much higher ratio than either the *Plowden* or Curtin violins. Also included in Fig. 8 was the measured directivity for the *Titian* and *Plowden*. The higher OP/IP ratio for the *Titian* was accompanied by higher directivity compared to the *Plowden*; the Curtin violin directivity was similar to the *Plowden* (both near average) and consistent with its top OP/IP ratio, and the assumption of a relatively constant ratio for the back. Moreover, an arching-

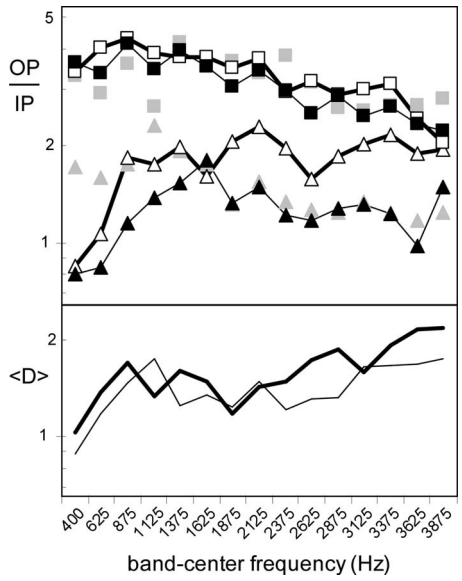


FIG. 8. Top panel: OP/IP (log) mobility ratio vs band-center frequency. Top plate—△ (*Titian*, open, thick line; *Plowden*, filled, thin line; Curtin, gray); back plate—□ (*Titian*, open, thick line); *Plowden*, filled, thin line; *Willemotte*, gray). Bottom panel: (log) directivity: *Titian* (thick line), *Plowden* (thin line).

directivity correlation, if real, offers a possible link to the OP/IP ratio depending on arching also. Of course, the flat plate (arch=zero) has only OP motion to first order, and thus a very high OP/IP ratio.

2. OP–IP, BH, and the bridge island

One poorly understood aspect of violin vibrations is the origin of the BH peak near 2.4 kHz in the OP mobility *and* accompanying radiativity spectra (Fig. 2). This structure seems unusual only because of its magnitude, with typical R_{eff} and ζ_{tot} values for its frequency region. Bridge rocking about the waist leading to up–down antiphase bridge feet motions that excite corpus OP motion have been proposed as a physical mechanism for this peak.[22,23] Examination of OP motion near 2.4 kHz, however, indicated little such antiphase motion at the bridge feet, although as noted a significant peak in OP motion was observed in $\langle Y_{\text{corpus}} \rangle$.

The BH peak has shown some sensitivity to changes in the bridge rocking mode frequency f_{rock}, especially at f_{rock} values closest to 2.4 kHz, where its amplitude and centroid frequency slumped noticeably (cf. Fig. 10, Ref. 16). This experimental observation links the BH peak with the bridge, but does not lead to an obvious explanation of the underlying mechanism. Understanding this region is crucial because string energy enters the corpus here through a tuned substructure whose optimized coupling to the corpus is essential to good violin sound[16] *and* because abundant nearby boundaries–discontinuities—especially so in the *X* direction—create a favorable environment for extensional →flexural transformations that can lead to acoustic radiation.

If bridge rocking motions do not produce the BH peak what does? Durup and Jansson, in a systematic "violin" experiment using a simplified geometry violin (flat rectangular plates), observed a BH peak only *after* cutting simplified *f*

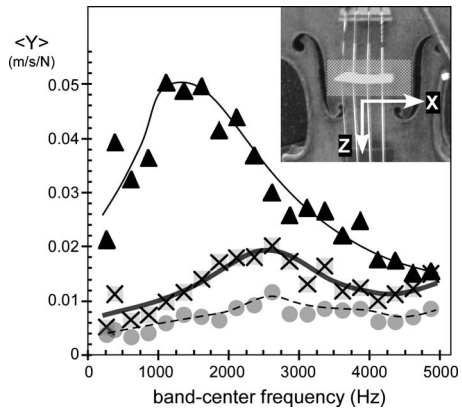


FIG. 9. Three-violin-average three-dimensional bridge-island rms mobilities vs band-center frequency (lines drawn to guide eye only): $\langle Y_Y \rangle$—▲ (thin solid line), $\langle Y_X \rangle$—☒ (thick solid line), $\langle Y_Z \rangle$—● (without Curtin violin- see the text; thin dashed line). Nominal intra-band s.d. $\approx \pm 0.007$ for $\langle Y_Y \rangle$ and $\langle Y_X \rangle$. (Inset: Hatched bridge-island region superimposed on violin photo with X–Z axes notated.)

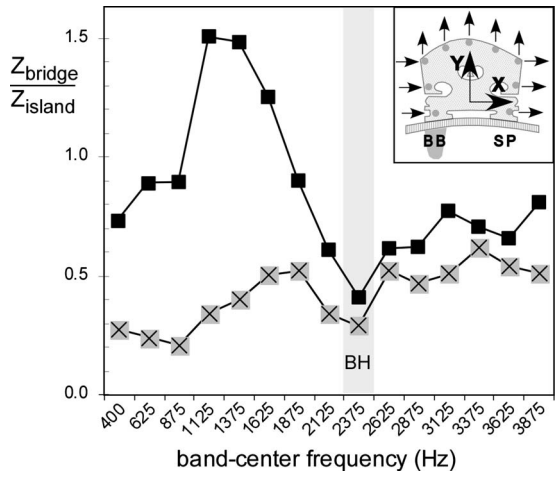


FIG. 10. Impedance ratio of bridge (1D) to bridge island (3D) vs band-center frequency for Y(■) and X(☒) directions. (Inset: Bridge point measurement location–direction).

holes (three-segment, squared-off, elongated "S" shapes) into the top plate. Crucially, only the long straight section led to the BH peak *and* no BH peak appeared when this section was above the bridge.[17] If cutting *f* holes is essential to creating both a bridge island and the BH peak, perhaps the simultaneous reduction in X direction stiffness around the bridge feet *and* creation of close-by boundaries–discontinuities offers a plausible augmenting mechanism for OP motion, viz. an alternative IP bridge→corpus →radiation path via extensional→flexural transformation with substantial subsequent radiation.

The 3D measurements also allowed us to investigate IP motions along X and Z directions directly. Mobilities were rms averaged over 19 points in a small region of the top plate near the bridge feet—a bridge island, here shown as an inset in Fig. 9 with X and Z axes noted. The island X-Y-Z mobility components (notated as $\langle Y_X \rangle$, $\langle Y_Y \rangle$, $\langle Y_Z \rangle$) for the *Plowden, Titian*, and Curtin violins were band-averaged separately, rather than having the X and Z components conglomerated into an overall IP motion as before.

The *Titian–Plowden*–Curtin three-violin-average bridge-island 3D mobility behaviors shown in Fig. 9 present some interesting differences from the corpus mobility (OP) results in Fig. 2:

(1) The $\langle Y_Y \rangle$ peak is now near 1.4 kHz, not 2.5 kHz as seen in Fig. 2, with a relatively smooth falloff above this.
(2) $\langle Y_Z \rangle$ is overall the lowest across the range, while $\langle Y_Y \rangle$ is the highest, likely reflecting relative stiffnesses along each direction, and possibly even the X-direction excitation at the bridge corner.
(3) $\langle Y_X \rangle$ has a definite, broad peak near 2.5 kHz, and was the only mobility component to show a definite peak in the BH region.
(4) $\langle Y_X \rangle$ for the *Plowden* was close in magnitude (~70%) to $\langle Y_Y \rangle$ near 2.4 kHz, with the *Titian* being much lower (~30%).
(5) From 625 to 4875 Hz the *Titian* $\langle Y_Y \rangle / \langle Y_X \rangle$ ratio is about twice the *Plowden* and Curtin values. It is quite possible this difference reflects the OP/IP corpus ratio and directivity results presented in Fig. 8.
(6) The Curtin violin results were quite similar to those of the *Plowden*, with comparable $\langle Y_Y \rangle$ (and $\langle Y_X \rangle$) magnitudes and trends. The $\langle Y_X \rangle$ peak was broader than for either old Italian; $\langle Y_Z \rangle$ was too weak-noisy for reliable analysis and was not included in Fig. 9.

The presence of the bridge-island $\langle Y_X \rangle$ peak near 2.5 kHz for the three-violin average suggests significant X-plane motion accompanying bridge rocking motion in string-corpus energy transfer. Unfortunately, there were no bridge 3D measurements. Accordingly the 12-violin 1D bridge measurements were reanalyzed to look at X and Y bridge motions separately, and then used in conjunction with 3D bridge-island results to compute a bridge/bridge-island impedance ratio.

The bridge 1D rms mobility results had previously shown a prominent BH peak when averaged over X and Y mobilities;[16] reanalysis of X (side-only) and Y (top-only) measurements (see Fig. 10 inset for point locations–directions) showed ~2.4 kHz peaks in rms mobility for both directions separately. Since string→bridge→bridge-island energy transfer is so important, a rough measure of direction-specific impedance relationships based on mobility average inverses was used to compute the X and Y impedance ratios, $Z_{\text{bridge}}/Z_{\text{island}}$, separately. These ratios, presented in Fig. 10, show a distinct local minimum in the BH region for X and Y directions, as well as a prominent maximum near 1.3 kHz for the Y direction. The trends seen in Figs. 9 and 10 are certainly suggestive of some bridge to bridge-island excitation mechanism based on X motion.

One additional, possibly pertinent note related to the old Italian CT scans was that all these violins showed some significant internal repair work around the bridge region where the soundpost and bridge feet (and bass bar replacement) tend over time to disrupt the structural integrity of the soft spruce top plate (Fig. 11). Localized repairs in this critical

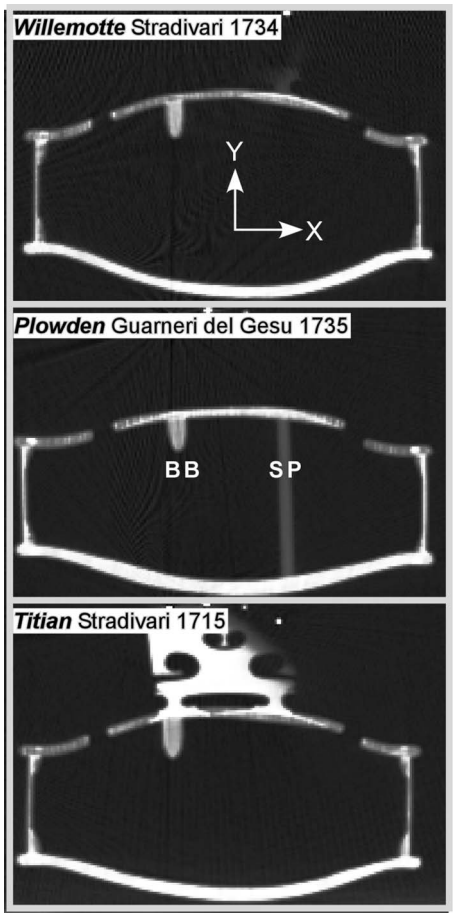


FIG. 11. CT scan slice in the bridge-soundpost region of three old Italian violins. Repairs under bridge feet in spruce top ($\rho \approx 0.4$ g/cm^3) over bass bar and soundpost are readily seen as higher-density (brighter) regions. (Maple back and bridge have $\rho \approx 0.6$ g/cm^3.)

energy transfer region are a natural companion to age and playing, and are universal to instruments of the classical period.

These repairs commonly require gluing in wood patches. If the patch density matched the original, and the fit were flawless only the (brighter) higher density glue-line arc would stand out in a CT scan. Standing out among these violins, the *Titian* also showed an additional, small high density patch underneath the bass bar-side bridge foot, as well as a prominent patch glue line under the soundpost-side bridge foot. Whatever the cause of the *Titian*'s high OP–IP ratio, these repairs are mentioned because of extensional $\rightarrow$flexural transformations possible at discontinuities. Perhaps density discontinuities at a patch glue line could be one such cause?

IV. CONCLUSIONS

Examining the totality of our experimental results for violins of widely varying quality it was impossible to reach a conclusion different than that of previous researchers, viz. that the very best violins *measure* little different from the worst (assuming all had been properly setup, of course). All had similar underlying structural acoustic behaviors: signature modes with unexceptional frequencies and total damping, total damping trends, radiation efficiency trends, fraction of vibrational energy turned into sound, etc., etc. What was observed were different spectral balances as the relative magnitudes of various psychoacoustically important regions changed over the profiles, perhaps tilting toward the low or high ends of the profile, or emphasizing a specific region. (Schleske has commented extensively on the effects of such shifts/emphases in spectral balance on perceived sound quality.[24])

The excellent violin radiativity profiles do reflect common remarks about the best violins—they are more "even" across the measured range, and strong in the lowest range. As to being loud? In a large auditorium where typical reverberation provides a low frequency boost and a high frequency rolloff above 5 kHz, the BH+bridge+F_{RAD} concentration of sound near 2–4 kHz (where the ear is most sensitive) combined with the higher frequency directivity "boost" seen for the *Titian* (but not the *Plowden* or *Willemotte*) could certainly help a solo violin being heard over the orchestra.

Can our experimental modal–acoustic results address such matters as pretreating wood with various chemicals, or deal with the varnish per se? Succinctly, no. Assuming constant shape, violin response to some driving force is determined by the overall stiffness-density properties of its various materials, irrespective of how these were arrived at. At this time the most productive area scientifically appears to be the radiativity profile, the "measureable" at the end of the energy trail that seems most immediate to a violinist in the overall judgment of violin sound quality. Structural acoustics modeling of the entire profile offers significant additional insight into quality-related differences.

Perhaps a contrarian viewpoint about quality might be useful here? What truly defines violin excellence? If the answer is truly excellent violinists, then the reliability-reproducibility of their psychoacoustic judgments must draw more attention. It would seem illogical to expect violinists who pride themselves on their personal sound not to prefer certain violins over others because they are better at creating that sound. If excellent violinists cannot agree on a quality rating because of sound preferences—or worse, rate two quite different sounding violins as good—shouldn't it follow that scientific measurements could do no better?

ACKNOWLEDGMENTS

The 3D measurements resulted from a cooperative effort among Polytec, Inc., which provided their advanced 3D laser system (operated by David Oliver and Vikrant Palant), the East Carolina University Acoustics Laboratory (acoustics scans, F.B. and Danial Rowe), renowned violin maker Samuel Zygmuntowicz, who arranged for the loan of the old Italian violins, the Violin Society of America (Joseph Regh and Fan-Chia Tao), which paid for the insurance and transportation of the violins, and Dr. Claudio Sibata at the CT scanner in the Leo Jenkins Cancer Center at East Carolina University. In addition, Samuel Zygmuntowicz and another renowned violin maker, Joseph Curtin, kept all instruments in optimum condition. All measurements used comprehensive test facilities constructed at East Carolina University between 1998 and 2003 with the aid of National Science

Foundation Grant No. DMR-9802656. Additionally, my former graduate students Machele Bailey, Ken Jacobs, Kuntao Ye, John Keiffer, and Jose Garcia-Cobian made important earlier contributions.

[1]K. D. Marshall, "Modal analysis of a violin," J. Acoust. Soc. Am. **77**, 695–709 (1985).
[2]G. Bissinger, "Some mechanical and acoustical consequences of the violin soundpost," J. Acoust. Soc. Am. **97**, 3154–3164 (1995).
[3]M. Schleske, "Eigenmodes of vibration in the working process of a violin," Catgut Acoust. Soc. J. **3**, 2–8 (1996).
[4]M. Schleske, "On making 'tonal copies' of a violin," Catgut Acoust. Soc. J. **3**, 18–28 (1996).
[5]M. Schleske, "Empirical tools in contemporary violin making. 1. Analysis of design, materials, varnish and normal modes," Catgut Acoust. Soc. J. **4**, 50–65 (2002).
[6]G. Bissinger and A. Gregorian, "Relating normal mode properties of violins to overall quality: Signature modes," Catgut Acoust. Soc. J. **4**, 37–45 (2003).
[7]G. Bissinger, "Modal analysis of a violin octet," J. Acoust. Soc. Am. **113**, 2105–2113 (2003).
[8]G. Bissinger and J. C. Keiffer, "Radiation damping, efficiency, and directivity for violin normal modes below 4 kHz," Acoust. Res. Lett. Online **4**, 7–12 (2003).
[9]G. Bissinger, "Contemporary generalized normal mode violin acoustics," Acust. Acta Acust. **90**, 590–599 (2004).
[10]G. Bissinger, E. G. Williams, and N. Valdivia, "Violin *f*-hole contribution to far-field radiation via patch near-field acoustical holography," J. Acoust. Soc. Am. **121**, 3899–3906 (2007).
[11]G. Bissinger, "The role of radiation damping in violin sound," ARLO **5**, 82–87 (2004).
[12]A. Runnemalm, N. E. Molin, and E. Jansson, "On operating deflection shapes of the violin body including in-plane motions," J. Acoust. Soc. Am. **107**, 3452–3459 (2000).
[13]G. Bissinger, "A unified materials-normal mode approach to violin acoustics," Acust. Acta Acust. **91**, 214–228 (2005).
[14]D. E. Oliver, V. Palan, G. Bissinger, and D. Rowe, "3-dimensional laser doppler vibration analysis of a stradivarius violin," *Proceedings of the 25th International Modal Analysis Conference, Society for Experimental Mechanics*, Bethel, CT, 2007, paper 372 (CD proceedings only).
[15]G. Bissinger, "A0 and A1 coupling, arching, rib height, and *f*-hole geometry dependence in the 2-degree-of-freedom network model of violin cavity modes," J. Acoust. Soc. Am. **104**, 3608–3615 (1998).
[16]G. Bissinger, "The violin bridge as filter," J. Acoust. Soc. Am. **120**, 482–491 (2006).
[17]F. Durup and E. Jansson, "The quest of the violin bridge-hill," Acust. Acta Acust. **91**, 206–213 (2005).
[18]E. Rohloff, "Ansprache der Geigenklänge (The speaking of violin sounds)," Z. Angew. Phy. **17**, 62–63 (1964). English abstract published in *Musical Acoustics, Part II, Violin Family Functions*, Benchmark Papers in Acoustics, Vol. **6**, edited by C. M. Hutchins (Dowden, Hutchinson & Ross, Stroudsburg, PA, 1976).
[19]C. M. Hutchins, "A 30-year experiment in the acoustical and musical development of violin-family instruments," J. Acoust. Soc. Am. **92**, 639–650 (1992).
[20]F. Fahy and P. Gardonio, *Sound and Structural Vibration: Radiation, Transmission and Response*, 2nd ed. (Academic, New York, 2007).
[21]L. Cremer, *The Physics of the Violin* (MIT, Cambridge, 1984).
[22]I. P. Beldie, "About the bridge hill mystery," Catgut Acoust. Soc. J. **4**, 9–13 (2003).
[23]J. Woodhouse, "On the bridge-hill of the violin," Acust. Acta Acust. **91**, 155–165 (2005).
[24]M. Schleske website, http://www.geigenforschung.de/11handbuch/en_extras3handbuch04klangfarbe.pdf. (last viewed 15 May, 2008).

Vibrational characteristics of harp soundboards

Chris Waltham[a] and Andrzej Kotlicki
Department of Physics and Astronomy, University of British Columbia, Vancouver British Columbia, V6 T 1Z1, Canada

(Received 25 January 2008; revised 24 April 2008; accepted 11 June 2008)

Harps exist in different forms, from large factory-made concert harps to small hand-made folk harps. This variety presents both a challenge and an opportunity for acousticians. The musical quality of a harp depends on many factors, but key among these is the soundboard. This work sets out to define some general desirable qualities of a harp soundboard. First, in order to understand the relationship between the vibrational behavior of a bare soundboard and that of a completed instrument, a 36-string harp was built from scratch. Measurements were made at each stage of construction, and the results showed how the bare soundboard properties affect those of the finished harp. Second, the soundboards of several harps of different sizes were assessed by measuring the admittances along the string bar. These data showed that one relationship crucial to the quality of the soundboard is that between the modal shapes and modal frequencies of the soundboard, and the position and fundamental frequencies of the strings attached to it. A general statement is made about this relationship, one which should be of use to harp makers.
[DOI: 10.1121/1.2956479]

PACS number(s): 43.75.Gh, 43.40.Dx [NHF] Pages: 1774–1780

I. INTRODUCTION

The classification system of Hornbostel and Sachs defines a harp to be a many-stringed chordophone whose soundboard is oriented perpendicular to the plane of the strings.[1,2] The harp is almost unique in that its soundboard is driven directly at the string attachment points, and not via a bridge. The attachment points are distributed along the length of the soundboard, generally along the centerline. Figure 1 shows the basic geometry. The string is plucked at a position that is generally somewhere between a half and a third of the distance from the soundboard to the neck. The string motion is nonplanar, and the soundboard is driven by the oscillating force vector from the string, which depends on both the varying polar and azimuthal angles it makes with the soundboard and also with the varying tension. These intrinsically nonlinear interactions between string and soundboard, and those between string and string via the soundboard, pose particular problems for the acoustician. The soundboard has its own set of resonances with different modal shapes.[3–5] Therefore its response to excitation by different strings can vary widely from one string to another. Unlike other instruments that exist in a single or a small set of highly evolved forms, modern harps come in many shapes and sizes, from small lap harps, through medium-size celtic-style harps,[6] to full-size concert instruments. In addition there are many historical styles that are still being used for performance. This means that acousticians interested in the harp must first look at general issues arising from its morphology rather than focus on a particular design.

In addition to these acoustical issues, there are many engineering difficulties associated with harp construction, the primary one being that the string tension has to be born directly by a soundboard, which is also the primary sound radiator. These two criteria oppose each other: mechanical stability requires a thick soundboard, while sound radiation requires a thin one. The harp is therefore an uneasy compromise, and good harps tend not to last for very long but merely for decades, as opposed to centuries for the violin family of instruments.

Beyond engineering issues, the harp manufacturer has to ensure that the string-to-string variation in sound radiation is kept to a minimum while the overall level of sound radiation is reasonably high. These qualities are controlled by the relationship between the modal shapes and frequencies of the soundboard, and the position and fundamental frequencies of the strings attached to it. The soundboard modes are determined predominantly by the shape and thickness of the soundboard and string bar, but also by the stiffness of the soundbox and the motion of air through the soundholes at the back. The progression of modal shapes of a bare soundboard is still recognizable when the soundboard is mounted on a soundbox in a completed harp, albeit with some modification caused by the soundbox and the tension imposed by the strings. This paper therefore concentrates on the soundboard, its geometry and the materials out of which it is made, and how these affect its vibrational behavior. The aim is to make the most general statements possible that may guide manufacturers of soundboards for any style of harp. It does not purport to be a general theory of harp acoustics. Other important issues not discussed in this paper include the air motion in the soundbox,[7] its radiativity,[8–10] and the strings.[11,12] There are also the interactions between string and soundboard, string and strings,[13] fingers and strings,[14] and the soundbox and performer.[15]

Section II of this paper discusses harp morphology and construction. Details are given for modern full-size concert harps, although most of these features exist in scaled-down

[a]Electronic mail: cew@phas.ubc.ca

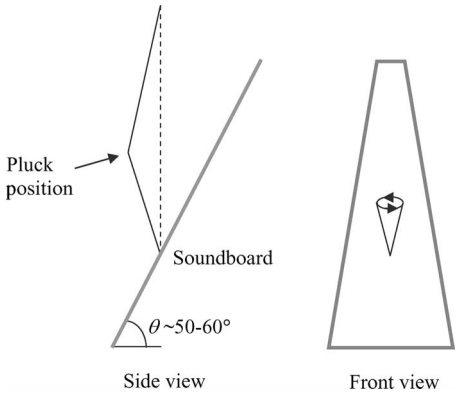


FIG. 1. Geometry of the connection between a harp string and the soundboard. The string is plucked typically just below the center of the string. The frontal view on the right shows the nonplanar motion of the string after plucking.

form in smaller harps. Section III is a summary of published material that describes how harps produce sound, and sets this present work on the vibrational behavior of soundboards in context. Section IV takes the reader through the construction of a soundboard and soundbox, and the measurements that were made at various junctures. Section V shows measurements on several harps of different types and manufacturers, and features common to all harps are discussed.

II. HARP CONSTRUCTION

A concert harp is basically a triangular structure, formed of the post or (fore) pillar, the neck, and a soundboard mounted on a soundbox. The strings are attached at one end to tuning pegs and bridge pins mounted in the neck and at the other end to the soundboard. The structure has to be made strong enough to withstand a total string tension of about 12–20 kN. The following specifications are taken from Salvi,[4,16] Lyon and Healy,[17] Camac,[18,19] and Aoyama[20] full-size concert harps. The soundboard is approximately trapezoidal in shape, around 1.4 m in length, 0.5 m wide at the base, 0.1 m at the top, and of thickness varying from 11–12 mm at the bottom (bass) end to 2–2.5 mm at the top (treble). It is made of strips of spruce (Sitka or Engelmann) between 3 and 8 cm wide bonded together and covered with a thin veneer, typically also of spruce. Sometimes, as in the case of the Salvi Aurora (Fig. 2), the soundboard has flared "wings" at the bass end, which protrude beyond the width of the soundbox. The strings are connected through two strong bars of hardwood mounted along the centerline of the soundboard, one on the outside (the cover bar visible on the front) and one heavier reinforcing bar inside. In addition, harmonic bars (made of beech or fir) are mounted approximately parallel to the string bars on the inside of the soundboard (Fig. 3).

The soundbox is semiconical in shape and is built up by bonding hardwood veneer (e.g., beech) around a mold. There are four or five soundholes in the back of the soundbox and one in the base. The primary function of these holes is to gain access to mount or replace strings, although they have important acoustical effects. Inside the back of the soundbox

FIG. 2. The configuration of a modern concert harp, a Salvi Aurora (Ref. 16). Author photo.

are strong U-shaped ribs (beech, aluminum, or steel), which prevent the box from undergoing too much flexure under the string tension.

Harps of all sizes and styles have more or less the same soundboard and soundbox morphology for reasons that will be outlined below, and the main variations are in size and aspect ratios. Generally harps are strung from the left side of the neck to the center of the soundboard. This creates an enormous torque on the neck and contributes to the short life span of strung harps. Some historical harps, particularly those of the Spanish baroque, have a second rank of chromatically tuned strings elsewhere on the soundboard.[1] Paraguayan harps are strung from a slot in the middle of the neck,[2] which is an obvious engineering solution to the torque problem.

A modern concert harp has 46 or 47 strings, running from C_1 or D_1 to G_7. Smaller harps like those involved in this study often have 36 strings, starting with B_1 or C_2. The lowest strings are mounted a few cm from the base of the soundboard, the highest strings a few centimeters from the top. The lower strings of concert harps are made of copper-wrapped steel, those in the midrange are gut, and the upper strings are nylon. Smaller harps use similar materials, except that the midrange strings are often nylon-wrapped nylon rather than (expensive) gut.

III. HARP ACOUSTICS

The classic approach to defining the vibroacoustic properties of a good instrument can be seen in the cases of well-

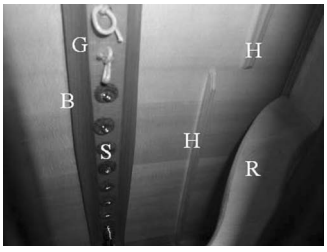


FIG. 3. View through one of the soundholes of a Salvi Aurora concert harp. Visible are the central reinforcing bar (B) at the division between gut (G) and steel strings (S), the horizontal strips of the soundboard, the vertical harmonic bars (H), and part of the soundbox reinforcement (R). The horizontal grain of the spruce can be seen in this view. Author photo.

studied instruments like the violin and guitar. For the violin, the great instruments from the masters of Cremona have been examined in great detail by physicists with respect to their acoustical properties, and sets of criteria have been developed for predicting the behavior of the completed instrument. With the availability of many fine instruments, it has been possible to define vibrational criteria for the critical individual components, in particular the top and back plates and the bridge.[21] A careful manufacturer can now follow these criteria and have a good chance of producing a competent instrument.[22]

Weinreich's analysis of the violin[23] has shown that it is very important to have the fundamental air resonance (A_0) and top plate resonance (T_1) well spaced in the low to mid range of the instrument. These two resonances contribute to the radiativity of the instrument with opposite signs in such a way as to guarantee that there are no dead notes, i.e., there are no frequencies where the radiativity is zero. All wooden shells with sound holes have these two types of resonances, and the harp is no exception. However, the harp sound holes are relatively large compared to those of a violin or guitar because their main function is access to string mounting points at the back of the soundboard. As a result, the air resonances are weak and hard to identify.[8] The relationship between air and soundboard motions have been studied by alternately blocking holes and adding chimneys to them.[24] These large but temporary changes allow identification, but their effects on the resonant structure of the whole are quite subtle. Hence one may expect that the general vibrational features of the bare soundboard remain identifiable once the soundboard is mounted onto the soundbox, although there will be frequency shifts. This is indeed what we find.

Two studies have been made of the sound radiation from concert harps. Bell and Firth[8] measured a Salvi Orchestra (a 46-string concert harp with an unflared trapezoidal soundboard) and found three distinct frequency regimes: below 400 Hz the harp is omnidirectional, but above this frequency the angular dependence of the radiation becomes progressively more complex. Le Carrou *et al.* studied the Camac Atlantide[10] (a 47-string concert harp) and attempted to model the radiation with one monopole for the soundboard, and either one or five monopoles for the five sound holes. They found the single monopole model to give reasonable results below 240 Hz, with some improvement by adding more for the sound holes. The phase of the radiation from these monopoles with respect to the soundboard excitation indicates that the T_1 mode lies below the A_0 mode.[24] Above this frequency the fit to one monopole was poor, and adding more monopoles yielded little improvement. A further study of the radiated power[9] showed that the sound from the sound holes dominated that from the soundboard in certain frequency bands, namely 150–200 and 400–550 Hz.

How the string excites the soundboard depends on the mobility of the soundboard at the string contact point and the frequencies of the string motion. How much sound is produced depends on the behavior of the rest of the soundboard, which, except at the bottom end of the range, does not act as a simple baffle. However, as we show, the geometry of the soundboard ensures that for any given resonance, there is always one dominant antinode, and this migrates up the soundboard as the frequency increases. This work is concerned with the relationship between the string attachment position and the modal shapes, in particular the positions of this primary antinode.

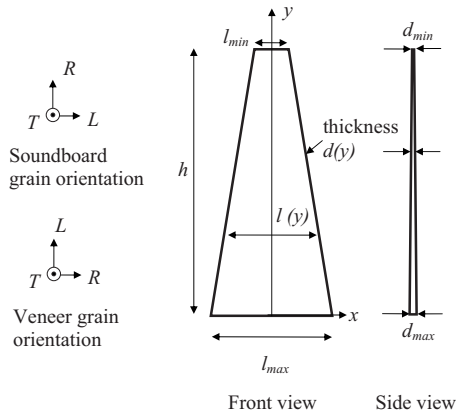


FIG. 4. Layout of harp soundboard, showing the *x*-*y* coordinate system used in this paper and the orientation of the wood grain for the soundboard base and veneer. The longitudinal (L), tangential (T), and radial (R) directions refer to the natural cylindrical coordinate system of a tree limb.

IV. MECHANICAL PROPERTIES OF HARP SOUNDBOARDS

The shape of a modern soundboard is a flat trapezoid with a linearly increasing thickness from top to bottom. The wood is typically spruce, and Fig. 4 shows the basic shape and orientations of the grain, which runs across its width (the *x*-direction). The soundboard consists of many pieces, a few centimeters wide, that are bonded together to form a sheet. The anisotropy of the spruce is high and so the board is much stiffer in the *x*-direction than in the *y*-direction.[25] The lower stiffness in the *y*-direction is necessary to give the right modal shapes, and the stiffer *x*-direction helps support the string tension. However, there is a tendency for the soundboard to distort and crack under the stress caused by the string tension, and this issue is addressed in two ways. Firstly a veneer of spruce is applied to the whole board,[26] with the grain running in the *y*-direction (Fig. 4). Secondly, strong string bars are applied to the front and back faces of the soundboard, drilled to allow the strings to pass through and be tied off at the back. We found that the string bars are also important in increasing the admittance of the upper end of the soundboard.

TABLE I. Inside dimensions of the bare soundboard when bonded to an open plywood frame. These are also dimensions of the finite-element model (see Fig. 4).

Shape	Trapezoid (mm)
Length h	983
Base width l_{max}	329
Base thickness d_{max}	4.7
Treble width l_{min}	66
Treble thickness d_{min}	2.4

TABLE II. Elastic properties of Sitka spruce (*P. sitchensis*). The values for moduli E_x and E_y and the density are experimental; the other values are taken from ratios given in *The Wood Handbook* (Ref. 25).

Parameter	Value for Sitka spruce
Elastic modulus E_x	12.3 ± 0.2 GPa
Elastic modulus E_y	0.89 ± 0.02 GPa
Elastic modulus E_z	0.53 GPa
Poisson ratio ν_{xy}	0.37
Poisson ratio ν_{yz}	0.43
Poisson ratio ν_{xz}	0.47
Shear modulus G_{xz}	0.79 GPa
Shear modulus G_{yz}	0.04 GPa
Shear modulus G_{xy}	0.75 GPa
Density ρ	441 ± 8 kg/m^3

A. Vibrational characteristics

The geometry of the soundboard produces a progression of modal shapes such that the fundamental mode of vibration has an antinode at the bass end, and the primary antinodes (the largest antinodes) of higher modes move progressively up the soundboard toward the treble end. This can be seen in results from the finite-element model that was created of a simple spruce trapezoid of dimensions given in Table I, with wood shown in Table II. The values for moduli E_x and E_y and the density are experimental, coming from tests of different strips used in the construction; the other values are taken from ratios given in *The Wood Handbook*.[25] Variability of a few percent from sample to sample is to be expected.[27] In the model, the edges were all clamped, as a first approximation to the way a soundboard is bonded to a soundbox. The first three modal shapes are shown in Fig. 5. Each mode has one antinode that is larger than all the others—this we call the primary antinode—that moves progressively up the soundboard as the frequency increases. Firth[3] and Firth and Bell[4] made a systematic study of the lowest modes of soundboards, varying the thicknesses and widths, and found these dimensions to be critical in maintaining these modal shapes. The relationship between frequency and mode number was shown to be approximately linear for typical soundboard shapes. The frequencies of these modes depend most critically on the value of the longitudinal elastic constant, E_x.

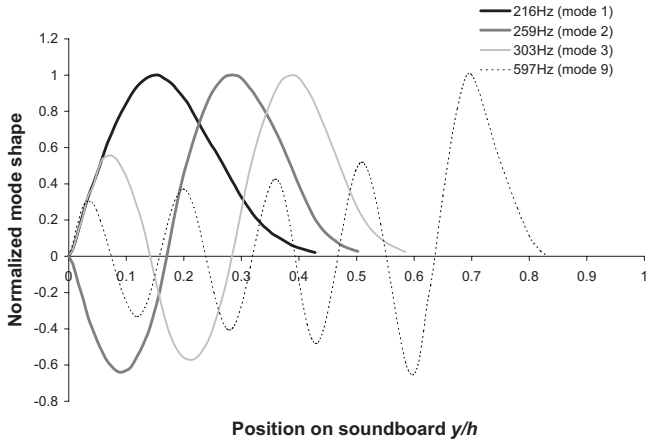


FIG. 5. Finite-element model predictions for the modal shapes for the first three (and ninth) modes of a trapezoidal spruce soundboard, clamped at all edges. The soundboard dimensions are as given in Table I and the spruce properties as in Table II. The modal shapes are normalized such that the primary antinodes have a unit amplitude.

B. Construction of a soundboard

For this work a soundboard of standard shape was created and measured at each stage of construction. The soundboard was ultimately destined for a replica of a 36-string 1820 Morley harp.[28] The interior dimensions (after subtracting the thickness of the soundbox) are given in Table I, wood properties in Table II, and exterior dimensions in Table III under "Morley," the harp for which it was intended. Strips of spruce 6–7 cm wide were cut and hand planed to within 10 μm of the required thickness. These were bonded edge to edge with polyvinyl acetate (PVA) adhesive (PVA, yellow carpenter's glue) and covered with a fir veneer of 0.75 mm thickness. The veneer was bonded to the spruce with a thin layer of hard hide glue. The string bars were made of maple 22 mm wide and 6 mm deep at the base and 13 mm wide and 2 mm thick at the top.

For the purposes of making preliminary measurements, the soundboard was temporarily mounted on an open frame made of a $\frac{3}{4}$ in. plywood. The frame was intended to approximate the way it would be supported in the soundbox, without the complications of air modes introduced by the enclosed air and sound holes. The soundboard on the frame was scanned by measuring the driving-point admittance at many points along the axis of symmetry, i.e., where the

TABLE III. Soundboard, soundbox, and soundhole dimensions (all millimeters, see Fig. 4 for definitions) for harps discussed in this paper. The lengths and widths are exterior values, the interior volume of the soundbox is V (in m^3), and the total area of the soundholes is A (in m^2). The soundboard thicknesses include the veneer (except the Salvi Orchestra).

Harp	Strings	h	$l_{\min}$	$l_{\max}$	$d_{\min}$	$d_{\max}$	V	A
Orchestra	46 (D_1–D_7)	1310	95	410	2	12	0.027	0.040
McFall	36 (C_2–C_7)	1083	101	360	7	9	0.02	0.022
Morley	36 (B_1–B_6)	1025	70	344	3.1	5.5	0.010	0.017

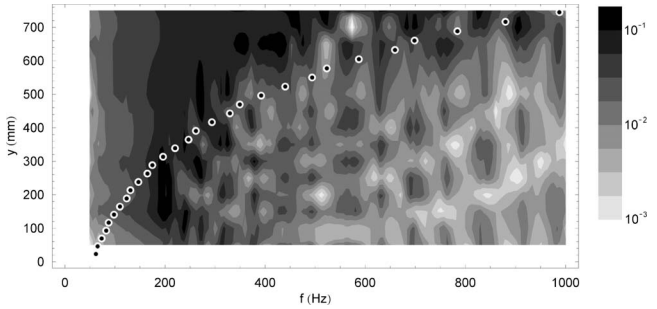


FIG. 6. Admittance data for the bare Morley harp soundboard (no veneer, no string bars, and no strings) temporarily mounted with all edges clamped to an open plywood frame. In this contour plot, the frequency is plotted as the horizontal axis, the position along the soundboard is the vertical axis, and the shading represents the admittance Y in s/kg (darker means higher). The points show the eventual position and fundamental frequency of the harp strings.

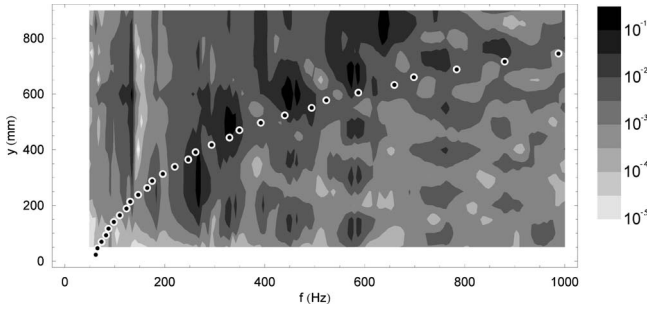


FIG. 7. Admittance data for the Morley harp soundboard, with veneer and string bars, mounted unstrung on the soundbox, plotted in the manner of Fig. 6 (note the admittance scale is different). The points show the position and fundamental frequencies of the harp strings.

strings would ultimately be attached. We make the assumption that the strings will primarily excite the symmetrical modes, ignoring for now second-order effects where nonplanar string motion excites twisting modes. The admittance was measured using a small, light (0.2 g) Endevco Isotron 25B accelerometer[29] and a tap hammer fabricated with a piezoelectric crystal. The accelerometer was calibrated by the manufacturer (and checked with a calibrated shaker). The tap hammer was calibrated by using it and the accelerometer to measure the inertance of a known mass.

The results for the admittance of the bare soundboard are shown in a contour plot (Fig. 6) in which the frequency is plotted as the horizontal axis, the position along the soundboard is the vertical axis, and the shading represents the admittance (darker means higher). The progression in modal shapes is very plain, and the regions of highest admittance move steadily up the soundboard as the frequency increases. Some splitting is observed in mode 4 (312–330 Hz) due to some imperfection of construction. The relationship between mode number and mode frequency is not precisely linear; the lower modes are somewhat closer together than those higher. The measured modal frequencies (198, 242, 274, 312/330, 378 Hz, etc.) show the same trend as those from a finite-element model of a single-piece soundboard clamped on all edges (216, 259, 303, 347, 393 Hz, etc.) but are in general 8% lower, indicating that the bonded multipiece test structure is a little more flexible than one would deduce from considering the properties of single pieces of spruce. Plotting the string fundamental frequency versus the attachment positions produces a "string trajectory" on the plot (marked by the points), which runs to the left of the fundamental antinode and below the higher primary antinodes. The veneer and string bars were now added. The increased mass and stiffness increased the fundamental frequency by a few hertz and spread out the modal patterns into the upper end of the soundboard. These effects are summarized at the end of the next section.

C. Soundbox construction

The shape of a harp soundbox is typically a truncated half-cone. For this work, the shell of the soundbox was constructed by bending a flexible 6 mm thick plywood sheet around a mold. The shell was stiffened with U-frames made of a strong clear $\frac{3}{4}$ in. birch ply. The shell was bonded onto end pieces, also made of $\frac{3}{4}$ in. birch ply. Four sound holes were cut in the shell and one in the base end piece. The whole was mated with the soundboard.

Once the soundbox was assembled, the vibrational behavior of the soundboard was again measured. The results are shown in Fig. 7, and the basic modal structure of the bare soundboard is still clearly visible. The more rigid structure and enclosed air has raised all the frequencies by about 100 Hz over the bare soundboard on the frame. The additional feature at 130–150 Hz is due to the soundbox and the stiffening ribs therein. Once the harp was strung, the string tension raised the lower modes by a further increment of about 30 Hz. The changes to the modal structure of the soundboard at each stage of the construction are summarized in Fig. 8. Note the effect of the veneer and string bar in moving the antinodes toward the upper part of the soundboard; for example, the primary antinode of the third mode has moved from $y/h \approx 0.5$ to $y/h \approx 0.7$.

V. COMPARING HARPS

The soundboxes of several concert harps and smaller "celtic-style" harps were measured for this study. The con-

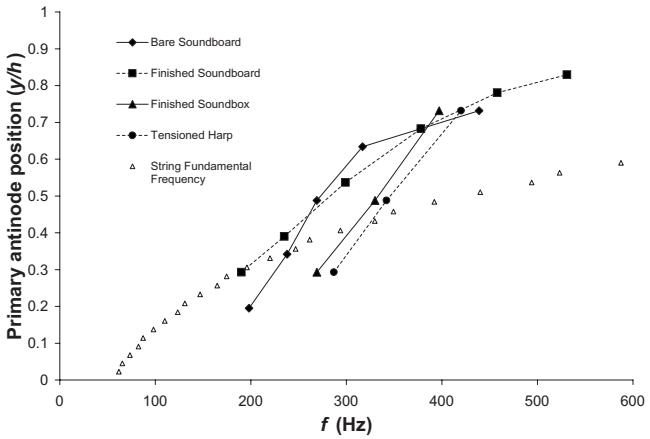


FIG. 8. Frequencies and primary antinode positions for a harp soundboard in four stages of construction: the bare soundboard, the finished soundboard with veneer and string bars, and the soundboard mounted in the soundbox, with and without tensioned strings. The triangles are the fundamental frequencies of the strings plotted against their attachment points.

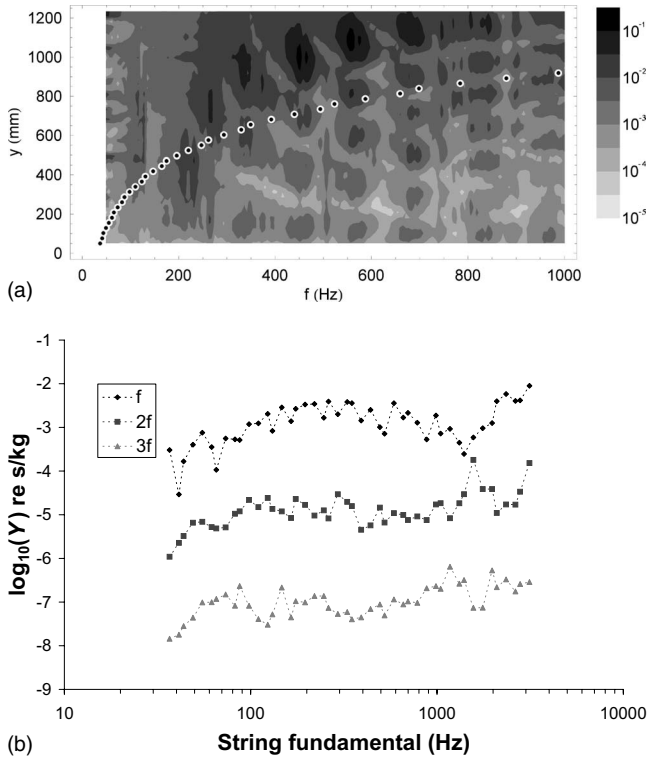


FIG. 9. Admittance data for the Salvi Orchestra. In the contour plot (a) the frequency is plotted as the horizontal axis, the position along the soundboard is the vertical axis, and the shading represents the admittance Y in s/kg (darker means higher). The graph (b) shows the admittance at each string attachment point at the frequencies of the fundamental and the next two overtones (shifted down by two and four units, respectively, for clarity).

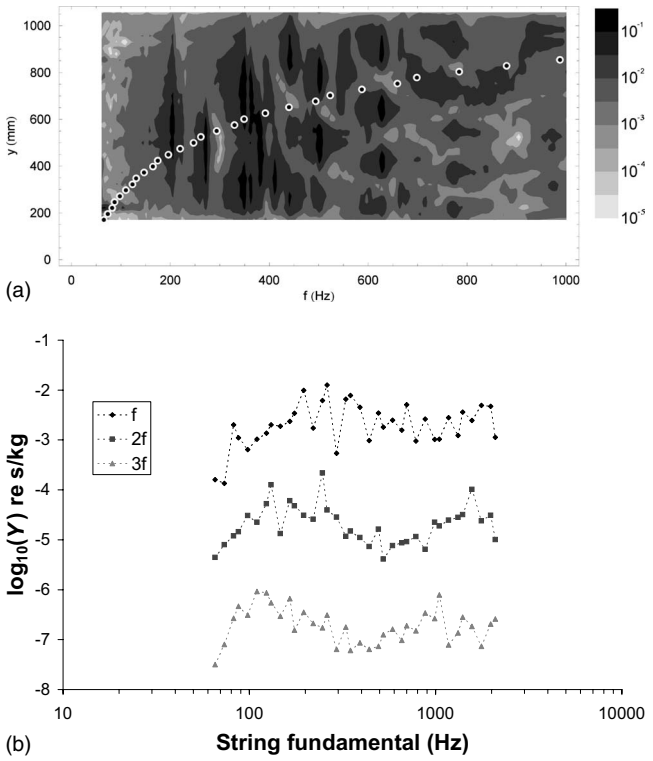


FIG. 10. Admittance data for the McFall, plotted in the same manner as for Fig. 9.

cert harps had very similar structures and a quality of sound that differed in details that are beyond the scope of this work; none had any significant vices and the admittance measurements produced very similar patterns. Data for the Salvi Orchestra are chosen for presentation here, as representative of the type, and because a study of this instrument has already been published by Firth and Bell.[4,8] The celtic-style harps were rather more variable in both geometry and admittance patterns. One was a replica of a harp made a century ago by James McFall of Belfast, and is chosen for examination here because of a large string-to-string variation in sound production.[30]

The soundboards were scanned as for the test soundboard by measuring the driving-point admittance at the base of each string on the completed, strung, and tensioned instrument. For the measurement, the strings were muffled. Dimensions of both instruments are given in Table III. The results are shown in contour plots (Figs. 9 and 10), which are plotted in the same manner as Figs. 6 and 7. The Salvi Orchestra, typical of the concert harps, shows in Fig. 9(a) clean modal patterns and obvious primary antinodes. The patterns are very similar to those seen in a bare soundboard, with evidence of some splitting. The regions of highest admittance move steadily up the soundboard as the frequency increases. For a given frequency these regions are mostly just above the string whose fundamental frequency corresponds to that frequency. The string trajectory runs below the primary antinodes, typically close to the secondary antinodes for all concert harps measured. Figure 9(b) is a graph of the admittance at each string attachment point at the frequencies of the fundamental and the next two overtones. What is seen here is a relatively smooth variation in admittance from string to string. However, this variation is not as small as it might be, as some of the strings coincide with the secondary antinodes. A small shift in the geometry or stringing could correct this. The situation is much less optimum for the McFall harp; here the modal patterns [Fig. 10(a)] are not as clean, and the regions of high admittance do not move toward the top of the soundboard as the frequency increases. The result is that the string trajectory passes through the regions of highest variability and the change in admittance from string to string, particularly in the center octave G_3 to G_4, is large. The high admittances correspond to strings which "boom" and the low admittances correspond to "dead" strings. Some of the variations at the fundamental frequency are also matched in the overtones, which is doubly unfortunate. For example, the 14th string (B_3) is high in both the fundamental and the first harmonic, and the 16th string is low in both the fundamental and the second harmonic. Both these strings are close together in the middle of the harp's range and therefore frequently played. A look at the dimensional specifications of the harps in Table III suggests a reason for the McFall's problem. The variation in soundboard thickness from top to bottom is very small—7 to 9 mm—compared to the other harps. This is the reason why the antinodes toward the top of the soundboard are not predominant: the soundboard is too stiff at the top end and too mobile at the bottom. It should be noted that this difficulty is not

intrinsic to the McFall design and arises from the soundboard construction of this particular model. This case is, however, instructive.

In comparing these two harps, one can see one aspect of why the Salvi Orchestra is a successful instrument. The trajectory of string positions and fundamental frequencies runs just above the antinode of the lowest mode and below the primary antinodes of the higher modes. This ensures good radiativity while avoiding large string-to-string variance. Why does the string trajectory not follow the primary antinodes more closely at the top and bottom ends? At the bottom there is a problem that the antinode of the fundamental mode never gets closer to the bottom than 25% of the length for the simple reason that this end of the soundboard is clamped to the soundbox and base of the harp. The bass strings are in any case primarily heard through their overtones and their excitation of higher strings.[13] The departure at the upper end arises from a compromise between radiativity of sound and alignment of frequencies. Getting upper strings to "sing" properly is an old problem in harp construction. To make the upper part of the soundboard resonate at frequencies closer to the string fundamental frequencies would require either thickening or narrowing the top of the soundboard. Either would increase the mechanical impedance of the upper end of the instrument and lower the sound radiation. The increased impedance would also make it even harder for primary antinodes to venture into that part of the board.

VI. CONCLUSION

Several harps large and small have been studied, in one case all the way through the construction process. It has been possible to track the vibrational behavior of the components of a harp soundboard through to that of completed and strung instruments. Detailed measurements on the completed instruments have been made, and these show a relationship between the patterns of resonant frequencies of the soundbox and the fundamental frequencies of the strings. Coincidences between string fundamental frequencies and the antinodes of the soundboard correlate with excessively loud strings; coincidences between string fundamental frequencies and the *nodes* of the soundboard correlate with dead strings. Armed with this information, it should be possible on occasion to improve unsatisfactory harps by shifting this relationship by changing the stringing. Where it is not possible to fix a given harp, one can still improve the next harp on the production line by changing the soundboard stiffness, by changing either its thickness or that of the string bar.

ACKNOWLEDGMENTS

We thank Domenic DiTomaso and David Axen for their technical assistance and the University of British Columbia School of Music for making their concert harps available. Many thanks also go to harpists Lori Pappajohn and Elizabeth Volpé-Bligh for sharing their instruments and for valuable discussions. The finite-element modeling was done by undergraduate students Laura Dunwoody, Tina Lee, Billy Lin, and Julia Lin.

[1]S. C. DeVale, "Harp," Grove Music Online, edited by L. Macy, www.grovemusic.com (last viewed April 18, 2008).
[2]R. Rensch, *Harps and Harpists* (Indiana University Press, Bloomington, IN, 1998).
[3]I. M. Firth, "Acoustics of the harp," Acustica **37**, 139–147 (1977).
[4]I. M. Firth and A. J. Bell, "On the acoustics of the concert harp's soundboard and soundbox," *Proceedings of the SMAC 83 Conference*, Stockholm, 1985, pp. 167–183.
[5]I. M. Firth, "Harps of the baroque period," Series II **1**, 52–61 (1989).
[6]It should be noted that modern celtic-style harps are basically scaled-down concert harps, without the pedal sharping mechanism. Medieval and Renaissance Irish harps are very different in structure and sound (see Refs. 1 and 2).
[7]A. J. Bell, "The Helmholtz resonance and higher air modes of the harp soundbox," **3**, 2–8 (1997).
[8]A. J. Bell and I. M. Firth, "The directivity of the concert harp," Acustica **69**, 26–30 (1989).
[9]F. Gautier and N. Dauchez, "Acoustic intensity measurement of the sound field radiated by a concert harp," Appl. Acoust. **65**, 1221–1231 (2004).
[10]J.-L. Le Carrou, F. Gautier, N. Dauchez, J. Gilbert, and J. R. de F. Arruda, "Low frequency model of the sound radiated by a concert harp," *Proceedings of Forum Acusticum*, Budapest, 2005.
[11]I. Firth, R. Bain, and A. Gallaher, "String design equations," **46**, 3–6 (1986).
[12]I. Firth, "Harp strings of the 18th century," Series II **1**, 19–23 (1988).
[13]J.-L. Le Carrou, F. Gautier, N. Dauchez, and J. Gilbert, "Modelling of sympathetic string vibrations," Acta Acust. Acust. **91**, 277–288 (2005).
[14]J.-L. Le Carrou, F. Gautier, F. Kerjan, and J. Gilbert, "The finger-string interaction in the concert harp," *Proceedings of the International Symposium on Musical Acoustics*, Barcelona, Spain, 2007.
[15]E. Panzer and J. Braasch, "The acoustic intensity and directivity of the modern concert harp," J. Acoust. Soc. Am. **121**, 3061 (2007).
[16]Salvi Harps, via Rossana 6, 12026 Piasco (CN), Italy.
[17]Lyon & Healy Harps Inc., 168 North Ogden Avenue, Chicago, IL 60607–1465.
[18]Les Harpes Camac, La Richerais BP 15, 44850 Mouzeil, France.
[19]J.-L. Le Carrou, "Vibro-Acoustique de la Harpe de Concert," Ph.D thesis, L'Université du Maine, Le Mans, France, 2006.
[20]Aoyama Musical Instrument Manufacturing Company, 15-62-1 Yoshinozakai Matsuoka-Cho, Yoshida-Gun, Fukui 910-1121, Japan.
[21]J. Alonso Moral and E. V. Jansson, "Eigenmodes, input admittance and the function of the violin," Acustica **50**, 329–337 (1982).
[22]E. Jansson, "Acoustics for Violin and Guitar Makers," www.speech.kth.se/music/acviguit4/ (last viewed April 18, 2008).
[23]G. Weinreich, "Sound hole sum rule and the dipole moment of the violin," J. Acoust. Soc. Am. **77**, 710–718 (1985).
[24]J.-L. Le Carrou, F. Gautier, and E. Foltête, "Experimental studies of the A_0 and T_1 modes of the concert harp," J. Acoust. Soc. Am. **121**, 559–567 (2007).
[25]D. W. Green, J. E. Winandy, and D. E. Kretschmann, *The Wood Handbook: Wood as an Engineering Material*, (Forest Products Society, Madison, WI, 2000), Chap. 4.
[26]I. M. Firth and A. S. Bell, "The acoustical effects of wood veneer," Acustica **66**, 114–116 (1988).
[27]V. Bocur, *Acoustics of Wood* (Springer, Berlin, 2006), p. 178.
[28]Plans from Robinson Harps of California, www.robinsonharp.com (last viewed April 16, 2008).
[29]Brüel and Kjaer, DK-2850 Nærum, Denmark.
[30]Plans from Robinson Harps, but these leave the soundboard thickness up to the builder.

Interference between wave modes may contribute to the apparent negative dispersion observed in cancellous bone

Christian C. Anderson, Karen R. Marutyan, and Mark R. Holland
Department of Physics, Washington University in St. Louis, St. Louis, Missouri 63130

Keith A. Wear
Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, Maryland 20993

James G. Miller[a)]
Department of Physics, Washington University in St. Louis, St. Louis, Missouri 63130

(Received 29 February 2008; revised 3 June 2008; accepted 5 June 2008)

Previous work has shown that ultrasonic waves propagating through cancellous bone often exhibit a linear-with-frequency attenuation coefficient, but a decrease in phase velocity with frequency (negative dispersion) that is inconsistent with the causality-imposed Kramers–Kronig relations. In the current study, interfering wave modes similar to those observed in bone are shown to potentially contribute to the observed negative dispersion. Biot theory, the modified Biot–Attenborogh model, and experimental results are used to aid in simulating multiple-mode wave propagation through cancellous bone. Simulations entail constructing individual wave modes exhibiting a positive dispersion using plausible velocities and amplitudes, and then summing the individual modes to create mixed-mode output wave forms. Results of the simulations indicate that mixed-mode wave forms can exhibit negative dispersion when analyzed conventionally under the assumption that only one wave is present, even when the individual interfering waves exhibit positive dispersions in accordance with the Kramers–Kronig relations. Furthermore, negative dispersion is observed when little or no visual evidence of interference exists in the time-domain data. Understanding the mechanisms responsible for the observed negative dispersion could aid in determining the true material properties of cancellous bone, as opposed to the apparent properties measured using conventional data analysis techniques.
[DOI: 10.1121/1.2953309]

PACS number(s): 43.80.Ev, 43.35.Cg, 43.35.Bf, 43.20.Hq [FD] Pages: 1781–1789

I. INTRODUCTION

There is considerable interest in the use of ultrasound to evaluate bone quality, with investigations of cancellous (trabecular) bone playing a prominent role. Cancellous bone consists of a network of (hard) calcified strands (trabeculae) through which courses (soft) bone marrow. Methods for characterizing bone tissue using ultrasound often consist of measurements of the velocity and attenuation properties of ultrasonic waves transmitted through the trabecular network.[1–5]

The intricate structure of cancellous bone tissue can complicate measurements. The material architecture is anisotropic, with the trabeculae predominantly oriented along the direction of stresses experienced by the bone. Consequently, acquired ultrasonic data depend on the angle of insonification relative to the predominant trabecular orientation.[6–8] A large number of ultrasonic investigations of cancellous bone reported in literature are performed on bovine leg bones or on human calcanei. In bovine leg bones, insonification is approximately perpendicular to the trabeculae in the medial-lateral (ML) and anterior-posterior directions, and approximately parallel in the superior-inferior (SI) direction. Human calcanei are typically insonified in the ML direction, corresponding to the perpendicular orientation.

The porous structure of cancellous bone supports the propagation of two compressional ultrasonic waves, often denoted as a fast wave and a slow wave.[6,9,10] Depending on experimental circumstances, such as the porosity and thickness of the bone sample and the insonification angle relative to the predominant trabecular orientation, the two waves are sometimes separate and visibly distinct in the time-domain radio frequency (rf) data; in other instances, only one wave is observed. One possible contributing factor for the observation of a single wave in certain cases is that the difference in the arrival times of the fast and slow waves is small compared to the temporal extent of the ultrasonic pulse, resulting in a rf trace that appears to be that of a single wave but is, in fact, the sum of interfering fast and slow waves. In their investigations of anisotropy in bovine cancellous bone, Hosokawa and Otani noted that the propagation speeds of the fast and slow waves converge as the incident ultrasonic field becomes aligned perpendicular to the predominant trabecular orientation, resulting in overlap of the fast and slow waves in

[a)]Electronic mail: james.g.miller@wustl.edu

the received rf trace.[6] Padilla and Laugier identified the overlap of fast and slow waves as a complicating factor in their study of a stratified model for bone.[9] Lee *et al.* reported that in the ML direction, fast and slow waves in bovine tibia overlap and "are observed as if one longitudinal wave propagates."[10] Haiat *et al.* also noted the difficulties associated with multiple-mode propagation.[11]

In general, investigators have found that the attenuation coefficient of bone varies linearly or quasilinearly with frequency, and typically report attenuation measurements as the rate of change in attenuation coefficient with frequency.[1,12] This parameter is known as broadband ultrasound attenuation or, if normalized by sample thickness, as normalized broadband ultrasound attenuation (nBUA), also known as *slope of attenuation*. However, despite the consensus on the frequency dependence of attenuation coefficient, there is considerable variation in measurements of the frequency dependence of phase velocity. Many laboratories report that on average the phase velocity of ultrasonic waves propagating through cancellous bone decreases with increasing frequency, a phenomenon known as negative dispersion.[1,2,4,13] However, an increase in phase velocity with frequency (positive dispersion) is observed in 10%–20% of investigated sites.[1,2]

The observed negative dispersion in the majority of bone samples analyzed is further confounded when one considers the apparent inconsistencies with the causality-imposed Kramers–Kronig (KK) relations that relate attenuation to dispersion.[12,14–17] According to the nearly local approximation to the KK relations with one subtraction, an *increase* in phase velocity with frequency is expected for samples exhibiting an approximately linear-with-frequency attenuation coefficient. If the attenuation coefficient varied strictly linearly with frequency, the dispersion curve would be positive and logarithmic.[12,15,16] The inconsistencies with the KK relations are especially troubling in light of evidence that they are valid even under conditions in which the attenuation coefficient and phase velocity exhibit complicated behavior.[18]

Currently, dispersion is not a widely used metric for clinical bone quality analysis. However, the negative dispersion measured in cancellous bone suggests that evaluating the material properties of bone using current data analysis techniques may be compromised because they result in the measurement of "apparent" material properties instead of the underlying "true" material properties. Consequently, a better understanding of the dispersion characteristics of cancellous bone, including the mechanisms responsible for the observed negative dispersion, could aid in determining the true material properties of cancellous bone, as opposed to the apparent properties measured using conventional data analysis techniques.

In a previous letter, our laboratory proposed that negative dispersion in cancellous bone can arise when rf signals consisting of overlapped fast and slow waves are analyzed as if they are a single longitudinal wave.[19] In the current investigation, we extend and enhance this proposed mechanism by using theoretical and experimental results of ultrasonic wave propagation in bone obtained by previous investigators as a basis to generate simulated fast and slow waves that are consistent with the nearly local approximation to the KK relations. We then use the simulated fast and slow waves to create a "mixed" wave form consisting of overlapping fast and slow waves, and then demonstrate that analyzing the mixed wave forms may contribute to the observed negative dispersion in cancellous bone.

II. THEORY

We considered theories of wave propagation in bone to aid in determining appropriate input values for the parameters used in our simulations. Ultrasonic wave velocity in bone has been extensively modeled using Biot theory,[7,8,20–28] modified Biot–Attenborough theory,[7,10,29,30] and stratified media theory.[7,9,31–34] Each of these models predicts the existence of fast and slow waves; however, the stratified model predicts that only the fast wave propagates at perpendicular insonification. Because our simulations involve two propagating compressional waves, we elected not to explore the stratified model in this study. We note, however, that actual bone structures are unlikely to be rigorously perpendicular to ultrasound wave propagation at any orientation, resulting in the presence of multiple modes even for nominally perpendicular orientations.

The Biot and modified Biot–Attenborough models are typically used to predict the phase velocities of fast and slow waves as functions of porosity, defined as (1 − volume fraction of bone). Each model has been empirically extended to include angle-dependent parameters to account for the anisotropy of cancellous bone.[7] Because the existing literature includes slightly different implementations of each theory, details of the theoretical formalisms we use to calculate phase velocities are included below.

A. Biot theory

Biot theory considers the motions of the fluid and solid components of a porous material. Input parameters consist of physical and mechanical properties of the fluid and solid components of the propagation medium. The input parameters are used to predict fast and slow compressional waves corresponding to in-phase and out-of-phase motions between the fluid and solid. Energy losses are due to viscous interactions at interfaces.

Biot theory gives rise to three elastic parameters P, Q, and R given by

$$P = \frac{\beta_{\mathrm{por}}\left(\frac{K_s}{K_f} - 1\right)K_b + \beta_{\mathrm{por}}^2 K_s + (1 - 2\beta_{\mathrm{por}})(K_s - K_b)}{D} + \frac{4\mu_b}{3}, \tag{1a}$$

$$Q = \frac{\left(1 - \beta_{\mathrm{por}} - \frac{K_b}{K_s}\right)\beta_{\mathrm{por}} K_s}{D}, \tag{1b}$$

$$R = \frac{K_s \beta_{\text{por}}^2}{D}, \tag{1c}$$

where $D = 1 - \beta_{\text{por}} - (K_b/K_s) + \beta_{\text{por}}(K_s/K_f)$ and K_s is the bulk modulus of the solid material, K_f is the bulk modulus of the fluid, K_b is the bulk modulus of the elastic frame, μ_b is the shear modulus of the elastic frame, and β_{por} is the porosity (volume fraction of fluid).

If the solid material is assumed to be isotropic, the bulk modulus K_s can be related to the intrinsic elastic parameters of the solid by

$$K_s = \frac{E_s}{3(1-2\nu_s)}, \tag{2}$$

where E_s and ν_s are Young's modulus and Poisson's ratio for the solid. Young's modulus of the elastic frame of the porous structure E_b is determined from the porosity and E_s through the power law relationship

$$E_b = E_s(1-\beta_{\text{por}})^n. \tag{3}$$

If the frame is assumed to be isotropic, its bulk and shear moduli K_b and μ_b can be written as

$$K_b = \frac{E_b}{3(1-2\nu_b)}, \tag{4}$$

$$\mu_b = \frac{E_b}{2(1+\nu_b)}, \tag{5}$$

where ν_b is Poisson's ratio for the frame.

Although Biot theory assumes an isotropic medium, it has been empirically extended to apply to anisotropic porous materials through exploitation of the power law relationship between E_b and E_s. The exponent n in Eq. (3) depends on the angle of insonification relative to trabecular alignment. Hosokawa and Otani found $n=1.46$ in the parallel orientation and $n=2.14$ in the perpendicular direction.[6] Williams found that n has a value of 1.23 when cancellous bone is insonified parallel to the trabeculae, and a value of 2.35 when insonification is perpendicular to the trabeculae.[26] Those values for n were adopted by Lee *et al.* to construct an angle-dependent Biot model.[7] In the current study, we elected to use the values established by Williams and later employed by Lee *et al.*, so we set $n=1.23$ and $n=2.35$ for the parallel and perpendicular directions, respectively.

Biot theory also includes mass coefficients to allow for viscous and inertial drag effects. These coefficients are given by

$$\rho_{11} + \rho_{12} = (1-\beta_{\text{por}})\rho_s, \tag{6a}$$

$$\rho_{22} + \rho_{12} = \beta_{\text{por}}\rho_f, \tag{6b}$$

$$\rho_{12} = -(\alpha_{\text{tort}}(\omega) - 1)\beta_{\text{por}}\rho_f, \tag{6c}$$

where ρ_f and ρ_s are the densities of the fluid and solid components and $\alpha_{\text{tort}}(\omega)$ is a structural factor known as the dynamic tortuosity. We have used the subscript tort to distinguish the tortuosity ($\alpha_{\text{tort}}(\omega)$) from the attenuation coefficient ($\alpha(\omega)$). The dynamic tortuosity was introduced by Johnson *et al.* as

$$\alpha_{\text{tort}}(\omega) = \alpha_\infty \left[1 + \frac{2}{\Lambda}\left(\frac{i\eta}{\omega\rho_f}\right)^{1/2} \right], \tag{7}$$

in which Λ is the viscous characteristic length and η is the fluid viscosity.[35] The tortuosity parameter α_∞ is given by

$$\alpha_\infty = 1 - s(1 - 1/\beta_{\text{por}}), \tag{8}$$

where s is a parameter derived from a microscopic model of a frame moving in a fluid. Other investigators have consistently let $s=0.25$.[6,7,25,26]

The elastic parameters and mass coefficients are used to construct a characteristic equation given by

$$\begin{vmatrix} \omega^2\rho_{11} - k^2 P & \omega^2\rho_{12} - k^2 Q \\ \omega^2\rho_{12} - k^2 Q & \omega^2\rho_{22} - k^2 R \end{vmatrix} = 0, \tag{9}$$

whose roots are the wave numbers of the fast and slow waves. The solutions are

$$k^2 = \omega^2\left(\frac{B \pm \sqrt{B^2 - 4AC}}{2A}\right), \tag{10}$$

where

$$A = PR - Q^2, \tag{11a}$$

$$B = R\rho_{11} + P\rho_{22} - 2Q\rho_{12}, \tag{11b}$$

$$C = \rho_{11}\rho_{22} - \rho_{12}^2. \tag{11c}$$

The phase velocities for the fast and slow waves are then calculated by $v_{\text{phase}} = \text{Re}(\omega/k)$, resulting in

$$v_{\text{fast}} = \text{Re}\left[\left(\frac{2A}{B - \sqrt{B^2 - 4AC}}\right)^{1/2}\right], \tag{12a}$$

$$v_{\text{slow}} = \text{Re}\left[\left(\frac{2A}{B + \sqrt{B^2 - 4AC}}\right)^{1/2}\right], \tag{12b}$$

where Re( ) returns the real component of a complex number.

B. Modified Biot–Attenborough model

The modified Biot–Attenborough model, proposed by Roh and Yoon, is a phenomenological approach for modeling wave propagation through a medium with cylindrical pores.[7,10,29,30] It allows for both viscous and thermal energy dissipations but requires empirically determined input parameters. We follow the formulation given by Lee *et al.*[7] The equation of continuity for one-dimensional wave propagation through a circular cylindrical pore is

$$-\rho_f \frac{\partial \langle v \rangle}{\partial x} = \frac{\partial \rho}{\partial t}, \tag{13}$$

where ρ_f is the fluid density and $\langle v \rangle$ is the average particle velocity over the cross section of the pore. The equation of motion is given by

$$\frac{\partial p}{\partial x} = \rho_c(\omega)\frac{\partial \langle v \rangle}{\partial t}, \tag{14}$$

where p is the acoustic pressure and $\rho_c(\omega)$ is the complex density given by

$$\rho_c(\omega) = \rho_f[1 - 2(\lambda e^{i\pi/2})^{-1}T'(\lambda e^{i\pi/2})]^{-1}, \tag{15}$$

where

$$T'(\lambda e^{i\pi/2}) = \frac{J_1(\lambda e^{i\pi/2})}{J_0(\lambda e^{i\pi/2})}, \tag{16}$$

in which J_0 and J_1 are the zeroth-order and first-order cylindrical Bessel functions. The dimensionless parameter λ is related to the size of the viscous boundary layer at the pore wall and may be written as

$$\lambda(\omega) = as_1(\omega/\nu)^{1/2}, \tag{17}$$

where a is the pore radius, ν is the kinematic viscosity of the fluid, and s_1 is a boundary condition parameter representing the pore frame rigidity.

When thermal effects are considered, the complex compressibility of the fluid $C_c(\omega)$ is given by

$$C_c(\omega) = (\gamma\rho_f c_f^2)^{-1}[1 + 2(\gamma - 1) \times (N_{\mathrm{Pr}}^{1/2}\lambda e^{i\pi/2})^{-1}T'(N_{\mathrm{Pr}}^{1/2}\lambda e^{i\pi/2})], \tag{18}$$

where c_f, γ, and N_{Pr} are the compressional speed, specific heat ratio, and Prandtl number of the fluid, respectively. When the model is extended to a material consisting of an ensemble of cylindrical pores, the wave number for the fast wave is given by

$$k_{\mathrm{fast}} = \alpha_{\mathrm{tort}}\left[\frac{k_c^2 k_s^2}{(1 - \beta_{\mathrm{por}})^{s_2}k_c^2 + \beta_{\mathrm{por}}^{s_2}k_s^2}\right]^{1/2}, \tag{19}$$

where α_{tort} is the tortuosity, β_{por} is the porosity, $k_s = \omega/c_s$ is the wave number of the pore frame, s_2 is a phase velocity parameter, and k_c is the complex frequency-dependent wave number of the pore fluid given by

$$k_c(\omega) = \omega[C_c(\omega)\rho_c(\omega)]^{1/2}. \tag{20}$$

The wave number for the slow wave is constructed in a similar fashion, under the assumption that the slow wave velocity tends to zero at low porosities. This wave number is given by

$$k_{\mathrm{slow}} = \alpha_{\mathrm{tort}}\left[\frac{k_c^2 k_g^2}{(1 - \beta_{\mathrm{por}})^{s_2}k_c^2 + \beta_{\mathrm{por}}^{s_2}k_g^2}\right]^{1/2}, \tag{21}$$

where $k_g = \omega/c_g$ is the wave number of a hypothetical fluid with an extremely low (approaching zero) wave velocity. The phase velocities for the fast and slow waves are found using $v_{\mathrm{phase}} = \mathrm{Re}(\omega/k)$, resulting in

$$v_{\mathrm{fast}} = \mathrm{Re}\left(\frac{\omega}{\alpha_{\mathrm{tort}}}\left[\frac{(1 - \beta_{\mathrm{por}})^{s_2}k_c^2 + \beta_{\mathrm{por}}^{s_2}k_s^2}{k_c^2 k_s^2}\right]^{1/2}\right), \tag{22a}$$

$$v_{\mathrm{slow}} = \mathrm{Re}\left(\frac{\omega}{\alpha_{\mathrm{tort}}}\left[\frac{(1 - \beta_{\mathrm{por}})^{s_2}k_c^2 + \beta_{\mathrm{por}}^{s_2}k_g^2}{k_c^2 k_g^2}\right]^{1/2}\right). \tag{22b}$$

The phase velocity parameter s_2 in Eqs. (19) and (21) determines the functional form of the phase velocity as a function of porosity. Lee *et al.* introduced an empirical anisotropy to this parameter and found that $s_2 = 0.5$ at parallel insonification and 1.7 at perpendicular insonification.[7] We adopted those values for s_2 in the current study.

TABLE I. Input model parameters used to generate the curves in Fig. 1.

Parameter	Biot	Modified Biot
Solid density (ρ_s)	1960 kg/m^3	1960 kg/m^3
Fluid density (ρ_f)	1000 kg/m^3	1000 kg/m^3
Young's modulus of solid (E_s)	20 GPa	
Bulk modulus of fluid (K_f)	2.28 GPa	
Poisson's ratio for solid (ν_s)	0.32	
Poisson's ratio for skeletal frame (ν_b)	0.32	
Tortuosity (α_{tort})	Equation (7)	1
Fluid viscosity (η)	10^{-3} Pa s	
Viscous characteristic length (Λ)	5 μm	
Exponent (n)	1.23 (parallel) 2.35 (perpendicular)	
Fluid compressional speed (c_f)		1500 m/s
Solid compressional speed (c_s)		3800 m/s
Kinematic viscosity of fluid (ν)		1×10^{-6} m^2/s
Specific heat ratio of fluid (γ)		1.004
Prandtl number of fluid (N_{Pr})		7
Pore radius (a)		0.5 mm
Boundary condition parameter (s_1)		1.5
Phase velocity parameter (s_2)		0.5 (parallel) 1.7 (perpendicular)

C. Biot and modified Biot–Attenborough model predictions

The input parameter values for each model are listed in Table I, and the resultant phase velocities at 500 kHz for the fast and slow waves as functions of porosity for both perpendicular and parallel insonifications are shown in Fig. 1.

As indicated in Fig. 1, the angle-dependent Biot and modified Biot models predict that when insonification is parallel to the trabecular orientation, corresponding to the SI direction in bovine leg bones, the velocities of the fast and slow waves remain distinct and moderately different over a wide range of porosities, including those within physiological ranges, and eventually converge toward the same value only at extremely high porosities. At a porosity of 0.85, Biot theory predicts that in the parallel orientation the difference in the velocities of the fast and slow waves is about 1180 m/s, whereas the modified Biot model predicts a difference of approximately 1320 m/s. Experimentally, the velocities at parallel insonification were sufficiently different to

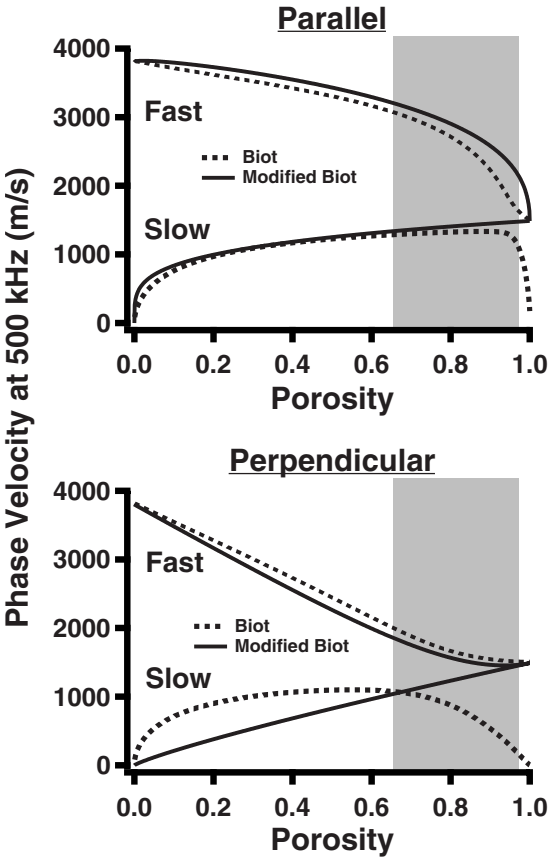


FIG. 1. Predictions of Biot theory (dashed lines) and modified Biot–Attenborough model (solid lines) for fast and slow wave velocities at 500 kHz as functions of porosity. The top panel shows predictions at parallel insonification, and the bottom panel shows predictions at perpendicular insonification. The shaded regions indicate the approximate range of physiological porosities for bovine and human cancellous bone measured by other investigators (Refs. 6, 10, and 25).

permit Hosokawa and Otani to observe and measure the velocities of distinct fast and slow waves at porosities above 0.8 in samples approximately 9 mm thick.[6]

The Biot and modified Biot model predictions differ for the porosity-dependent behavior of the velocities when bone is insonified in the clinically relevant perpendicular direction. In that orientation, and at a porosity of 0.85, Biot theory predicts a velocity difference of 865 m/s, whereas the modified Biot model predicts a difference of about 200 m/s. Experiments performed while insonifying in the perpendicular orientation have generally not resulted in the observation of distinct fast and slow waves. Investigations of the anisotropy of cancellous bone have demonstrated that the fast and slow waves become more and more overlapped as the angle of insonification proceeds from parallel to perpendicular.[6]

III. SIMULATED WAVE PROPAGATION IN BONE

We simulated the propagation of ultrasound through bone by generating fast waves and slow waves independently using phase velocity and attenuation coefficient parameters, and then combined the two waves to form a simulated received signal. The input to the simulation consisted of a simulated broadband reference pulse with a center frequency of 550 kHz and a −6 dB bandwidth of approximately

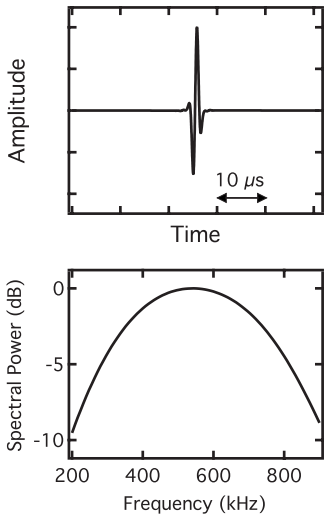


FIG. 2. The time-domain representation of the artificially generated reference pulse used as input to the simulations is displayed in the upper panel, and the power spectrum of the pulse is shown in the lower panel.

250–850 kHz (see Fig. 2). An output wave form corresponding to ultrasonic wave propagation through bone was produced by applying linear transfer functions to the input

$$\text{output}(\omega) = \text{input}(\omega)[H_{\text{fast}}(\omega) + H_{\text{slow}}(\omega)], \tag{23}$$

where output(ω) and input(ω) are the complex frequency-domain representations of the input reference wave form and output mixed wave form. We have assumed that wave propagation through bone and bonelike materials is linear, with speeds of sound and attenuation coefficients independent of amplitude. $H_{\text{fast}}(\omega)$ and $H_{\text{slow}}(\omega)$ are the transfer functions for the individual fast and slow waves given by

$$H_{\text{fast}}(\omega) = A_{\text{fast}} \exp[-\alpha_{\text{fast}}(\omega)l]\exp\left[i\omega\frac{l}{v_{\text{fast}}(\omega)}\right], \tag{24a}$$

$$H_{\text{slow}}(\omega) = A_{\text{slow}} \exp[-\alpha_{\text{slow}}(\omega)l]\exp\left[i\omega\frac{l}{v_{\text{slow}}(\omega)}\right], \tag{24b}$$

in which A_{fast} and A_{slow} are frequency-independent amplitude compensation factors that correspond to relative initial amplitudes of the fast and slow waves upon encountering the sample, l is the sample thickness, and $\alpha_{\text{fast}}(\omega)$ and $\alpha_{\text{slow}}(\omega)$ are the attenuation coefficients given by

$$\alpha_{\text{fast}}(\omega) = \beta_{\text{fast}}\frac{\omega}{2\pi}, \tag{25a}$$

$$\alpha_{\text{slow}}(\omega) = \beta_{\text{slow}}\frac{\omega}{2\pi}, \tag{25b}$$

where β_{fast} and β_{slow} are the values for nBUA. The phase velocities $v_{\text{fast}}(\omega)$ and $v_{\text{slow}}(\omega)$ are given by

$$v_{\text{fast}}(\omega) = v_{\text{fast}}(\omega_0) + v_{\text{fast}}(\omega_0)^2\frac{\beta_{\text{fast}}}{\pi^2}\ln\left(\frac{\omega}{\omega_0}\right), \tag{26a}$$

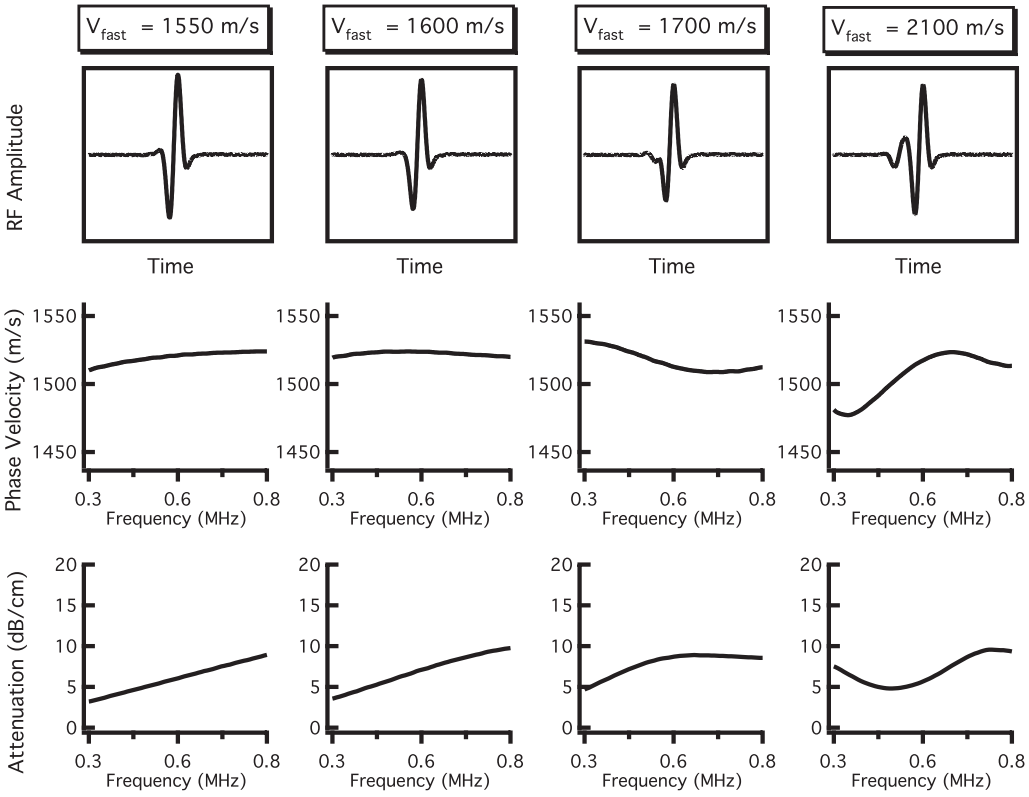


FIG. 3. Results of independent fast and slow wave propagations when the velocity of the fast wave is varied. The velocity of the fast wave increases from the left column to the right column. The top panels display the resultant mixed rf wave forms, and the center and bottom panels show the corresponding dispersion curves and attenuation coefficients obtained when the mixed wave form is analyzed as if it contained only one wave. When v_{fast} is only 50 m/s greater than v_{slow} (far left panels), the mixed wave form exhibits positive dispersion and a nearly linear attenuation coefficient. As v_{fast} becomes increasingly greater than v_{slow} (middle panels), the dispersion becomes negative. When v_{fast} is significantly larger than v_{slow} (far right panels), the dispersion curve and attenuation coefficient have complicated frequency-dependent behavior.

$$v_{\text{slow}}(\omega) = v_{\text{slow}}(\omega_0) + v_{\text{slow}}(\omega_0)^2 \frac{\beta_{\text{slow}}}{\pi^2} \ln\left(\frac{\omega}{\omega_0}\right), \qquad (26b)$$

where ω_0 is a chosen reference frequency of interest. The linear-with-frequency functional form of the attenuation coefficients [Eqs. (25a) and (25b)] is taken from the consensus of the published literature. The phase velocities [Eqs. (26a) and (26b)] are obtained by applying the nearly local form of the KK relations to the linear-with-frequency attenuation coefficients [Eqs. (25a) and (25b)], under the assumption that this form of the KK relations is valid for the fast and slow waves. These expressions for the phase velocities are valid for small dispersions.

In all simulations, the parameter l was set at 1 cm, a typical thickness of bone samples investigated *in vitro*. The value for the reference frequency $\omega_0/2\pi$ was set to 300 kHz, and $v_{\text{slow}}(\omega_0)$ was held constant at 1500 m/s, a velocity near that of sound in water. Based on measurements by Waters and Hoffmeister, the input values for β_{fast} and β_{slow} were set to 20 and 6.9 dB/cm MHz, respectively.[12] Input values for $v_{\text{fast}}(\omega_0)$, A_{fast}, and A_{slow} were systematically varied. Based on the theoretical predictions displayed in Fig. 1, and the experimental observations of overlapping fast and slow wave modes, we varied $v_{\text{fast}}(\omega_0)$ between 1550 and 2100 m/s, covering a wide range of velocities that result in the overlap of the fast and slow waves. When v_{fast} was varied, A_{fast} and A_{slow} were held constant at 0.3 and 0.7, respectively, so that the relative amplitudes of the fast and slow waves were comparable to those observed by other investigators.[6,12] Determining appropriate values for A_{fast} and A_{slow} is challenging, given the complexity involved in calculating transmission coefficients specific to individual fast and slow waves. We therefore created a second set of simulations in which $v_{\text{fast}}(\omega_0)$ was held constant at 1600 m/s and A_{fast} and A_{slow} were varied between 0 and 1 to determine the effects of those parameters on the resultant mixed wave form.

Phase velocities were calculated using a phase unwrapping algorithm previously shown to be valid for experimental data.[36]

IV. RESULTS

The resulting attenuation coefficients and phase velocities when $v_{\text{fast}}(\omega_0)$ was varied between 1550 and 2100 m/s are shown in Fig. 3. The relative contributions of the fast and slow waves, A_{fast} and A_{slow}, were held constant at 0.3 and 0.7, respectively. The left column of Fig. 3 was created with $v_{\text{fast}}(\omega_0)$ set at 1550 m/s, a velocity only 50 m/s faster than that of the slow wave. Under such conditions, a visual inspection of the rf signal reveals no explicit evidence of the presence of a second wave; furthermore, the attenuation coefficient and phase velocities do not exhibit suspicious behavior, despite interference taking place in the simulated ultrasonic field. However, as the difference in v_{fast} and v_{slow}

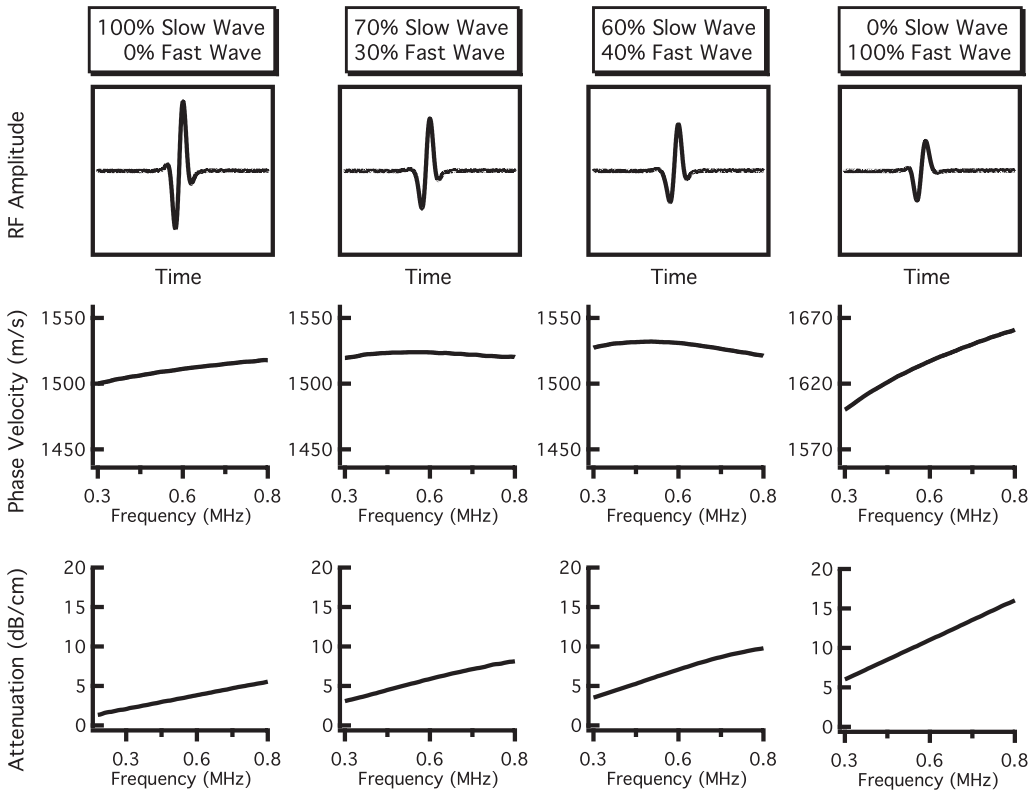


FIG. 4. Results of independent fast and slow wave propagations when the relative contributions of the fast and slow waves are varied. The contribution of the fast wave increases from the left column to the right column. The top panels display the resultant mixed rf wave forms, and the center and bottom panels show the corresponding dispersion curves and attenuation coefficients obtained when the mixed wave form is analyzed as if it contained only one wave. The far left-hand panels display only slow wave propagation ($A_{\text{fast}}=0$, $A_{\text{slow}}=1$), resulting in a positive dispersion and a strictly linear attenuation coefficient. The far right-hand panels show only fast wave propagation ($A_{\text{fast}}=1$, $A_{\text{slow}}=0$), again corresponding to a positive dispersion and linear attenuation coefficient. The middle panels represent mixed wave forms (nonzero amplitudes for both fast and slow waves) that exhibit negative dispersions while maintaining an approximately linear attenuation coefficient.

becomes more pronounced, as in the remaining columns of Fig. 3, a negative dispersion becomes apparent, in spite of the fact that the frequency-dependent behavior of the attenuation coefficient could be approximated by a linear-with-frequency fit. When $v_{\text{fast}}(\omega_0)=1600$ m/s, the magnitude of dispersion (difference in phase velocity) for a twofold increase in frequency is approximately −4 m/s, based on a linear least squares fit over the bandwidth from 400 to 800 kHz, as in the second column of Fig. 3. When $v_{\text{fast}}(\omega_0)=1700$ m/s, the dispersion magnitude is approximately −15 m/s over the same bandwidth, with a local minimum in phase velocity appearing around 650 kHz. Additionally, the mixed wave form begins to show some evidence of interference under visual inspection. By the time the fast wave velocity reaches 2100 m/s (right panel), the fast and slow waves are sufficiently separated in time that the dispersion curve obtained by analyzing the mixed wave form using a conventional phase spectroscopy algorithm has large fluctuations, with analogous behavior appearing in the attenuation coefficient.

Results for varying A_{fast} and A_{slow} are shown in Fig. 4. In the leftmost panel, the only contribution to the mixed wave form is the slow wave, and the resulting attenuation coefficient and phase velocity corresponding to this wave form are representative of the slow wave input parameters with positive dispersion. As the contribution of the fast wave increases (that is, A_{fast} becomes proportionally more important compared to A_{slow}), the behavior of the phase velocity changes and begins to decrease with frequency (demonstrates a negative dispersion) over portions of the bandwidth. When $(A_{\text{fast}},A_{\text{slow}})=(0.7,0.3)$, the magnitude of dispersion is −4 m/s over the 400–800 kHz bandwidth, but the attenuation coefficient can still be described as reasonably linear. The dispersion becomes approximately −11 m/s over the same bandwidth when $(A_{\text{fast}},A_{\text{slow}})=(0.6,0.4)$, with the attenuation coefficient beginning to exhibit a modest degree of not-linear-with-frequency behavior. In each case, the individual fast and slow waves each exhibit a logarithmically increasing (positive) dispersion and a linear-with-frequency attenuation coefficient. Furthermore, the time-domain rf data (top panels) do not show explicit evidence of two-wave interference upon visual inspection. The far right column of Fig. 4 shows the resultant mixed wave and its properties when $(A_{\text{fast}},A_{\text{slow}})=(1,0)$, corresponding to only fast wave propagation. As expected, the phase velocity and attenuation coefficient of the mixed wave form return to mimicking those of the fast wave with positive dispersion.

V. DISCUSSION

This study uses a linear system approach, in conjunction with established experimental and theoretical evidence for multiple-wave propagation in cancellous bone, to generate a phenomenological model capable of producing simulated

signals that exhibit negative dispersions similar to those observed in cancellous bone. The frequency-domain linear system model has the advantage of simplicity and ease of computation, but this heuristic approach may not adequately account for all features of ultrasonic propagation through cancellous bone. Alternative approaches, including those that take advantage of finite difference time-domain techniques, may prove better for relating the fast and slow waves to the material properties of bone.

We have examined the parameter space for this model, specifically the relative amplitudes and velocities of the fast and slow waves, in the limited manner outlined in Figs. 3 and 4. The dispersions become more extreme, and more negative, as the fast and slow waves become closer in magnitude and more disparate in velocity. The detailed nature of the dispersion, however, depends on complicated interdependencies among all the parameters in the model, and a rigorous determination of what regions of the parameter space generate negative dispersion is beyond the scope of this preliminary study.

We have assumed that the attenuation coefficients of cancellous bone rise linearly with frequency. This approximation appears to be adequate for frequencies in the hundreds of kilohertz range currently employed in many clinical devices and is consistent with the general consensus in the published literature. However, this model may not be rigorously valid for bandwidths extending far into the megahertz range, especially if longitudinal-to-longitudinal scattering becomes a significant contributor to the attenuation coefficient at higher frequencies.

The primary focus of this study is simulated propagation of multiple longitudinal waves through cancellous bone. Experimental confirmation of the results presented here would presumably involve analyzing data from bone samples that exhibit a negative dispersion when analyzed under the assumption that one wave was present and recovering two waves when the data are analyzed in a way that permits distinguishing between fast and slow waves. However, differentiating the fast and slow waves can be difficult when they are strongly overlapped temporally. Our laboratory has proposed a Bayesian approach for recovering the properties of each wave and has experimentally confirmed that negative dispersion can arise from two-mode propagation in a simple phantom.[37,38] A better understanding of the physical properties of cancellous bone that contribute to the attenuation and velocity properties of the fast and slow waves may also prove advantageous in predicting the degree of overlap of the two signals, which in turn could provide insight about the dispersion properties of a signal analyzed as if only one wave were present. To that end, Haiat *et al.* investigated the effects of anisotropy and bone volume fraction on the degree of separation between the fast and slow waves.[11]

VI. CONCLUSION

Simulated ultrasonic wave propagation of fast and slow waves with phase velocities and attenuation coefficients, similar to those predicted and observed in bone, yielded resultant wave forms consisting of interfering waves that exhibited a decrease in phase velocity with frequency when analyzed conventionally under the assumption that only one wave was present. The underlying "true" ultrasonic characteristics of the fast and slow waves were obscured when analyzed in such a manner. Although the individual wave modes exhibited strictly linear-with-frequency attenuation coefficients and positive logarithmic-with-frequency phase velocities, conventional analysis of the mixed wave forms frequently resulted in complex frequency dependencies of attenuation coefficient and dispersion. In fact, the two-independent mode model used in these simulations produced resultant mixed wave forms characterized by negative dispersion despite the fact that the individual fast and slow waves exhibit positive dispersions in accordance with the Kramers–Kronig relations for attenuation and dispersion. In some simulations, negative dispersion was produced without substantially changing the coarse visual characteristics of either the time-domain signal or the linear-with-frequency behavior of the attenuation coefficient. Although further study is required to determine the validity and role of independent mode simulations for characterizing bone, the interference of multiple waves appears to be a possible source for an apparent negative dispersion of the kind observed in cancellous bone. If the material and structural properties of cancellous bone give rise to multiple interfering wave modes, recovering the ultrasonic characteristics of the individual wave modes, instead of those of the mixed wave form, could provide more robust ultrasonic determinations of bone quality.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grant No. CBET-0717830 Scholar in Residence at the FDA and by NIH Grant No. R37 HL40302.

[1]P. Droin, G. Berger, and P. Laugier, "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592 (1998).
[2]K. A. Wear, "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646 (2000).
[3]B. K. Hoffmeister, S. A. Whitten, and J. Y. Rho, "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone (N.Y.) **26**, 635–642 (2000).
[4]K. A. Wear, "Group velocity, phase velocity, and dispersion in human calcaneus in vivo," J. Acoust. Soc. Am. **121**, 2431–2437 (2007).
[5]K. N. Apostolopoulos and D. D. Deligianni, "Influence of microarchitecture alterations on ultrasonic backscattering in an experimental simulation of bovine cancellous bone aging," J. Acoust. Soc. Am. **123**, 1179–1187 (2008).
[6]A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).
[7]K. I. Lee, E. R. Hughes, V. F. Humphrey, T. G. Leighton, and M. J. Choi, "Empirical angle-dependent biot and mba models for acoustic anisotropy in cancellous bone," Phys. Med. Biol. **52**, 59–73 (2007).
[8]E. R. Hughes, T. G. Leighton, P. R. White, and G. W. Petley, "Investigation of an anisotropic tortuosity in a Biot model of ultrasonic propagation in cancellous bone," J. Acoust. Soc. Am. **121**, 568–574 (2007).
[9]F. Padilla and P. Laugier, "Phase and group velocities of fast and slow compressional waves in trabecular bone," J. Acoust. Soc. Am. **108**, 1949–1952 (2000).
[10]K. I. Lee, H.-S. Roh, and S. W. Yoon, "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293 (2003).
[11]G. Haiat, F. Padilla, F. Peyrin, and P. Laugier, "Fast wave ultrasonic propagation in trabecular bone: Numerical study of the influence of porosity and structural anisotropy," J. Acoust. Soc. Am. **123**, 1694–1705 (2008).

[12]K. R. Waters and B. K. Hoffmeister, "Kramers-Kronig analysis of attenuation and dispersion in trabecular bone," J. Acoust. Soc. Am. **118**, 3912–3920 (2005).
[13]P. H. Nicholson, G. Lowet, C. M. Langton, J. Dequeker, and G. Van der Perre, "A comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurement in trabecular bone," Phys. Med. Biol. **41**, 2421–2435 (1996).
[14]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[15]K. Waters, M. Hughes, J. Mobley, G. Brandenburger, and J. Miller, "On the applicability of Kramers-Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[16]K. R. Waters, J. Mobley, and J. G. Miller, "Causality-imposed (Kramers-Kronig) relationships between attenuation and dispersion," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 822–833 (2005).
[17]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).
[18]A. Q. Bauer, K. R. Marutyan, M. R. Holland, and J. G. Miller, "Is the Kramers-Kronig relationship between ultrasonic attenuation and dispersion maintained in the presence of apparent losses due to phase cancellation?," J. Acoust. Soc. Am. **122**, 222–228 (2007).
[19]K. R. Marutyan, M. R. Holland, and J. G. Miller, "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–EL61 (2006).
[20]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[21]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[22]T. Haire and C. Langton, "Biot theory: A review of its application to ultrasound propagation through cancellous bone," Bone (N.Y.) **24**, 291–295 (1999).
[23]Z. E. A. Fellah, J. Y. Chapelon, S. Berger, W. Lauriks, and C. Depollier, "Ultrasonic wave propagation in human cancellous bone: Application of Biot theory," J. Acoust. Soc. Am. **116**, 61–73 (2004).
[24]N. Sebaa, Z. E. A. Fellah, M. Fellah, E. Ogam, A. Wirgin, F. G. Mitri, C. Depollier, and W. Lauriks, "Ultrasonic characterization of human cancellous bone using the Biot theory: Inverse problem," J. Acoust. Soc. Am. **120**, 1816–1824 (2006).
[25]K. A. Wear, A. Laib, A. P. Stuber, and J. C. Reynolds, "Comparison of measurements of phase velocity in human calcaneus to Biot theory," J. Acoust. Soc. Am. **117**, 3319–3324 (2005).
[26]J. L. Williams, "Ultrasonic wave propagation in cancellous and cortical bone: Prediction of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112 (1992).
[27]A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).
[28]A. Chakraborty, "Prediction of negative dispersion by a nonlocal poroelastic theory," J. Acoust. Soc. Am. **123**, 56–67 (2008).
[29]H.-S. Roh and S. W. Yoon, "Acoustic diagnosis for porous medium with circular cylindrical pores," J. Acoust. Soc. Am. **115**, 1114–1124 (2004).
[30]K. I. Lee, V. F. Humphrey, T. G. Leighton, and S. W. Yoon, "Predictions of the modified Biot-Attenborough model for the dependence of phase velocity on porosity in cancellous bone," Ultrasonics **46**, 323–330 (2007).
[31]E. R. Hughes, T. G. Leighton, G. W. Petley, and P. R. White, "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821 (1999).
[32]K. A. Wear, "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083 (2001).
[33]W. Lin, Y. X. Qin, and C. Rubin, "Ultrasonic wave propagation in trabecular bone predicted by the stratified model," Ann. Biomed. Eng. **29**, 781–790 (2001).
[34]T. J. Plona, K. W. Winkler, and M. Schoenberg, "Acoustic waves in alternating fluid/solid layers," J. Acoust. Soc. Am. **81**, 1227–1234 (1987).
[35]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[36]R. L. Trousil, K. R. Waters, and J. G. Miller, "Experimental validation of the use of Kramers-Kronig relations to eliminate the phase sheet ambiguity in broadband phase spectroscopy," J. Acoust. Soc. Am. **109**, 2236–2244 (2001).
[37]K. R. Marutyan, G. L. Bretthorst, and J. G. Miller, "Bayesian estimation of the underlying bone properties from mixed fast and slow mode ultrasonic signals," J. Acoust. Soc. Am. **121**, EL8–EL15 (2007).
[38]A. Q. Bauer, K. R. Marutyan, M. R. Holland, and J. G. Miller, "Negative dispersion in bone: The role of interference in measurements of the apparent phase velocity of two temporally overlapping signals," J. Acoust. Soc. Am. **123**, 2407–2414 (2008).

Direct methods for characterizing high-intensity focused ultrasound transducers using acoustic streaming

Matthew R. Myers,[a] Prasanna Hariharan, and Rupak K. Banerjee
Division of Solid and Fluid Mechanics, Center for Devices and Radiological Health, U. S. FDA, 10903 New Hampshire Avenue, Building 62, Room 2231, Silver Spring, Maryland 20993-0002

(Received 22 February 2008; revised 16 June 2008; accepted 18 June 2008)

Two techniques are presented for noninvasively determining the intensity field of high-intensity focused ultrasound transducers in a liquid medium. The techniques are based upon the streaming velocity induced in the liquid by the absorbed ultrasound beam. The approaches are similar to an iterative streaming method previously reported, but the present approaches are "direct:" The differential operations of the Navier–Stokes equations are performed directly upon the experimentally measured streaming velocity, rather than through an iterative approach that minimizes the difference between a theoretical estimate of the streaming velocity and the one measured experimentally. As such, the direct methods are much faster than the iterative technique. The price paid for the increase in speed is smaller spatial coverage; the direct techniques are applicable only where accurate streaming velocity is available. Comparisons performed in the range 100–1000 W/cm^2 focal intensity showed differences between the direct methods and the iterative streaming technique to be less than 20%. Similar differences were observed in low-power comparisons with hydrophone measurements.
[DOI: 10.1121/1.2957937]

PACS number(s): 43.80.Ev, 43.35.Yb, 43.25.Nm [DLM] Pages: 1790–1802

I. INTRODUCTION

In the safety evaluation of high-intensity focused ultrasound (HIFU) transducers, an important first step is the determination of the intensity field in a liquid medium. Direct measurement of the pressure or intensity at energy levels used in clinical applications can be difficult, due to the possibility of sensor damage or interference of the sensor with the HIFU beam (Shaw and ter Haar, 2006). Hence, it is worth pursuing noninvasive methods for characterizing HIFU transducers.

In a previous paper (Hariharan *et al.*, 2008), it was shown that acoustic streaming may be employed to characterize transducers without perturbing the acoustic field. The steady flow field induced by the absorbed ultrasound was measured using digital particle image velocimetry (DPIV). The acoustic intensity giving rise to the measured flow field was then determined through an iterative computational procedure involving repeated solution of the governing equations of ultrasound beam propagation (KZK equation) and fluid flow (Navier–Stokes). Pressure predictions from the iterative streaming technique closely matched (within 10%) hydrophone measurements at low power, where hydrophones can be safely used. A drawback of the iterative streaming technique is the extensive computation time required to repeatedly solve the governing equations during the optimization phase.

An alternative to the iterative streaming method is a "direct streaming method," wherein the velocity field derived from DPIV measurements is inserted directly into the equations governing the streaming motion, in order to determine the intensity contained in the source term. A difficulty with direct streaming calculations is the inherent instability of numerical differentiation of experimental data. In computing a derivative via finite differencing, for example, dividing by the small distance between adjacent velocity values tends to accentuate experimental uncertainties. This magnification of error increases further when second derivatives (e.g., those describing the viscous stress) are computed.

In this paper we present two direct streaming methods for determining the ultrasound intensity field. The first, the "Gaussian curve fitting" (GCF) method, fits the streaming velocity data to an analytic expression involving Gaussian functions and their integrals. Errors associated with differentiation of the experimental data are reduced by analytically differentiating the closed-form expressions. Our second technique, the finite difference (FD) method, uses finite differences to compute the derivatives contained in the Navier–Stokes operators. Errors are controlled by determining an "optimal" grid spacing—one large enough that experimental uncertainties are not excessively magnified, and small enough that extreme averaging of the field does not occur. If desired, curve fitting of differentiated data can be employed for additional smoothing.

Relative to the iterative approach in Hariharan *et al.* (2008), the direct techniques are much faster. Execution times for the direct methods were just seconds on a standard personal computer. The direct methods also require no software for simulating beam propagation or fluid flow. As a result, the direct methods can be cheaper to employ. A drawback of the direct methods is that they can generate reliable intensity estimates only in locations where accurate velocity

[a]Author to whom correspondence should be addressed. Electronic mail: matthew.myers@fda.hhs.gov

data are available. With the iterative method, a complete description of the acoustic intensity field is obtained.

The following section describes the direct methods. In Sec. III the methods are applied to sets of experimental data for three different transducers. The experiments were performed using the DPIV method described in Hariharan *et al.* (2008). Comparisons are made with the iterative approach and with experimental hydrophone measurements. In Sec. IV, the advantages and disadvantages of the direct methods are discussed.

II. METHODS

We consider an axisymmetric HIFU transducer radiating into a liquid medium of Newtonian viscosity μ, ultrasound absorption α, density ρ_0, and speed of sound c_0. We denote the direction along the axis of symmetry of the transducer the z axis, with the r-coordinate perpendicular to z. The liquid medium may be any Newtonian fluid, such as water, though it was found previously (Hariharan *et al.*, 2008) that hydrodynamic instabilities at higher intensities can be avoided by using a more viscous medium. In the experiments cited in this paper the streaming fluid will always be a Natrasol-based medium having a Newtonian viscosity of 12 centipoise and an ultrasound attenuation of 0.00055 Np/cm (Hariharan *et al.*, 2008).

The HIFU intensity is assumed to be in the "moderate" regime—large enough to be clinically relevant yet low enough that the beam propagation may be adequately described by linear acoustics (Hariharan *et al.*, 2008). In terms of focal intensity, the upper limit of the range of interest is around 1000 W/cm^2. Transducer characterization at higher intensities is discussed briefly in Sec. IV.

The steady fluid flow, i.e., the acoustic streaming field, induced by the absorbed ultrasound energy is assumed known from experimental measurements. Our velocity measurements were made using DPIV, though any method possessing sufficient resolution to accurately capture the velocity field in the focal region would suffice. A typical speed contour derived from our DPIV measurements is shown in Fig. 1. The contour represents the magnitude of the streaming velocity in the focal region of the ultrasound beam. Due to the high intensity, the focal region is the location of primary interest, and the one to which we will confine our attention in this paper. From Fig. 1 it can be seen that the extent of the focal region is a few cm in the axial direction and a few mm radially.

To derive the equation relating the acoustic intensity to the measured streaming velocity, we denote the fluid velocity generated by the ultrasound beam u_j. In the most general case, the indices i, j may take on the values 1,2,3 for the different Cartesian coordinate directions. The streaming velocity is $\overline{u_j}$, with the overbar denoting time averaging. The equation of motion for the streaming velocity is (Lighthill, 1978):

$$\rho_0\left(\overline{-u_i\frac{\partial u_j}{\partial x_i}}\right) = F_j - \frac{\partial \bar{p}}{\partial x_j} + \mu\nabla^2\overline{u_j}, \tag{2.1}$$

where

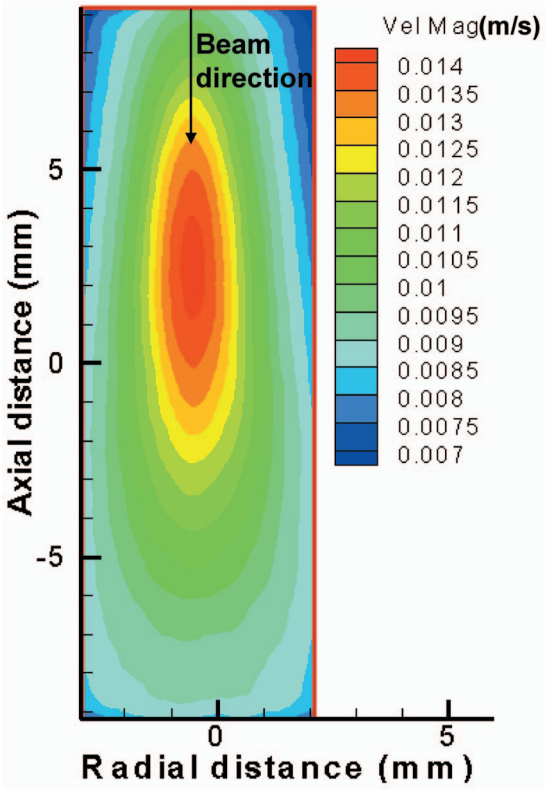


FIG. 1. Streaming velocity contour for HIFU-3 transducer. Acoustic power: 5 W.

$$F_j = -\frac{\partial(\overline{\rho u_i u_j})}{\partial x_i} \tag{2.2}$$

is the force per unit volume arising from the transfer of momentum from the acoustic field to the streaming fluid. As noted by Lighthill (1978), F_j is also the spatial variation of the Reynolds stress $\overline{\rho u_i u_j}$.

In the focal region, the HIFU beam may be modeled as a column of locally planar waves traveling parallel to the z axis. In that case the radiation force (per unit volume) on the fluid may be written (Nyborg, 1965)

$$F_z = \frac{2\alpha}{c_0} I_z, \tag{2.3}$$

where I_z is the z component of the time-averaged ultrasound intensity. Upon taking the z component of Eq. (2.1), incorporating Eq. (2.3), and solving for I_z in cylindrical coordinates, we obtain

$$I_z = \frac{c_0}{2\alpha}\left[\rho_0\left(\overline{-u_z\frac{\partial u_z}{\partial z}} + \overline{-u_r\frac{\partial u_z}{\partial r}}\right) + \frac{\partial \bar{p}}{\partial z} - \mu\nabla^2\overline{u_z}\right]. \tag{2.4}$$

In streaming flows located far from boundaries, the mean pressure gradient is typically small and can be neglected (Wu and Du, 1993). Also, because the radial component of velocity u_r is small compared to the axial component of velocity u_z in the focal region, and because radial derivatives are constrained to be zero on the z axis, the approximation

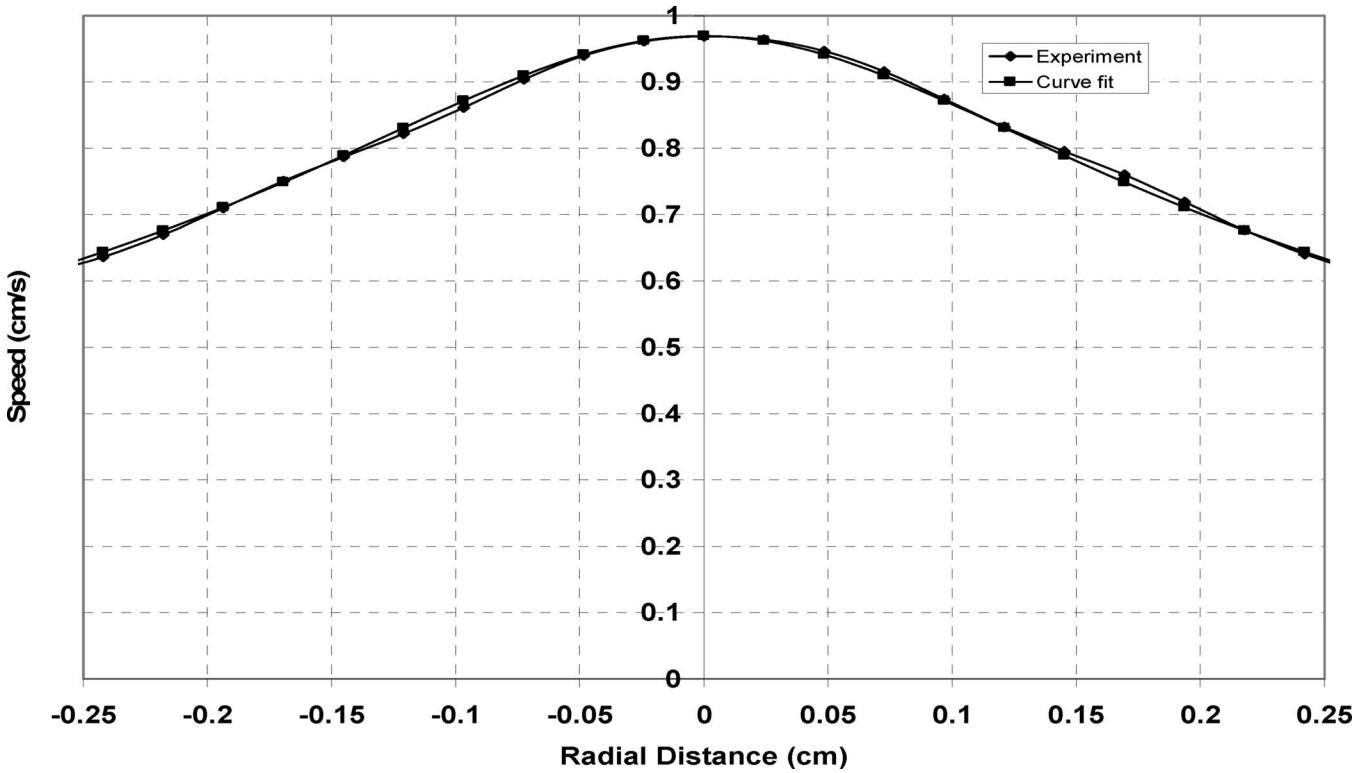


FIG. 2. Comparison of radial streaming velocity profile obtained from (i) PIV experiments and from (ii) curve fitting the experimental velocity with the error function expression. Transducer: HIFU-2; acoustic power: 3.6 W.

$$\overline{u_r}\frac{\partial \overline{u_z}}{\partial r} \approx 0 \tag{2.5}$$

can be made in the focal region. With these simplifications, the axial component of Eq. (2.4) becomes

$$I_z = \frac{c_0}{2\alpha}\left[\rho_0\left(\overline{u_z}\frac{\partial \overline{u_z}}{\partial z}\right) - \mu\nabla^2\overline{u_z}\right]. \tag{2.6}$$

In Hariharan *et al.* (2008), the intensity was found in an iterative fashion by refining the estimate for the intensity field that gave rise to the experimentally observed velocity field. In the direct approach we perform the operations on the right side of Eq. (2.6) directly upon the velocity field determined experimentally. Because the differentiated velocity field is less smooth than the velocity (and the twice-differentiated field even less so), care must be taken in order to produce intensity estimates with acceptable levels of accuracy. Below we present two techniques for performing the differential operations and smoothing the results.

A. Gaussian curve fitting (GCF) method

If the various derivatives contained in Eq. (2.6) are extracted from the finite-element calculations performed in the iterative method of Hariharan *et al.* (2008), it can be seen that the second radial derivative possesses an approximately Gaussian shape as a function of radial position. We therefore assume a profile of the form

$$\frac{\partial^2 u_z}{\partial r^2} = A(z) + C(z)\exp(-B^2(z)r^2), \tag{2.7}$$

where the functions $A(z)$, $B(z)$, and $C(z)$ are to be determined. Upon integrating Eq. (2.7) twice with respect to r and enforcing $\partial u_z/\partial r = 0$ on $r=0$, we obtain

$$u_z = \frac{1}{2}Ar^2 + \frac{Cr\sqrt{\pi}\,\mathrm{erf}(Br)}{2B} + \frac{C\exp(-B^2r^2)}{2B^2}, \tag{2.8}$$

where erf is the error function (Abramowitz and Stegun, 1972). The coefficients A, B, and C are determined for any desired value of z by fitting the form (2.8) to the streaming velocity data at the desired axial location. Fitting was performed using a least-squares method (Mathworks, Inc., 2002). A fit of the streaming velocity to the functional form of Eq. (2.8) is shown in Fig. 2, for transducer HIFU-2 operating at a power level of 3.6 W.

Radial derivatives required by Eq. (2.6) could be performed by analytically differentiating Eq. (2.8). Axial derivatives were computed by performing finite-difference operations on the velocity profiles in Eq. (2.8), since the coefficient functions $A(z)$, $B(z)$, and $C(z)$ were not described analytically. The finite-differencing procedure is described in the following section.

B. Finite difference (FD) method

The derivatives in Eq. (2.6) can be approximated using a variety of finite-difference (FD) formulas (Gerald, 1978). Examples are the following central-difference approximations to the first and second derivatives:

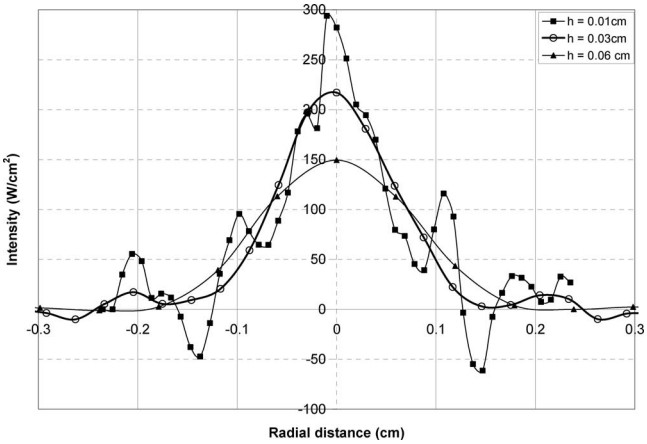


FIG. 3. Effect of mesh size: Acoustic intensity (W/cm^2) as a function of radial distance (cm) obtained from finite-difference method for three different mesh sizes (i) $h=0.01$ cm, (ii) 0.03 cm, and (iii) 0.06 cm. Acoustic power: 5 W; transducer: HIFU-3.

$$f'(x) = \frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h}, \tag{2.9}$$

$$f''(x) = \frac{-f(x+2h) + 16f(x+h) - 30f(x) + 16f(x-h) - f(x-2h)}{12h^2}. \tag{2.10}$$

For the FD method, we used Eqs. (2.9) and (2.10) to compute the derivatives of Eq. (2.8) in both the radial and axial directions. The minimum step size h we considered was the grid spacing over which the results from the DPIV method were obtained, approximately 0.1 mm. However, this led to a rather jagged intensity profile, as shown in Fig. 3. As h was increased, the profile became smoother, as indicated by the curve for $h=0.3$ mm in Fig. 3. Eventually, increases in h caused excessive averaging of the velocity profile. Using $h=0.6$ mm resulted in an intensity profile that was 30% lower on axis than the profile for $h=0.3$ mm, and absent of any secondary maxima (Fig. 3). Best agreement between finite-difference results and those obtained using the iterative technique of Hariharan *et al.* (2008) was obtained using a step size in the range of 0.3–0.4 mm. Subsequent finite-difference results presented in this paper are based upon $h=0.3$ mm.

III. RESULTS

The streaming experiments mentioned in Sec. II, and discussed in more detail in Hariharan *et al.* (2008), were performed using the three HIFU transducers featured in Table I. All experiments were performed in the 12 cP Natrosol medium, using the DPIV protocol described in Hariharan *et al.* (2008). Intensity fields for the three transducers were derived from the streaming velocity fields using the GCF and FD direct methods. Comparisons with hydrophone measurements and the iterative method were performed to evaluate the accuracy of the direct methods. Most of the experiments were performed at low power to enable hydrophone measurements.

TABLE I. Physical characteristics of HIFU transducers used in the experiments.

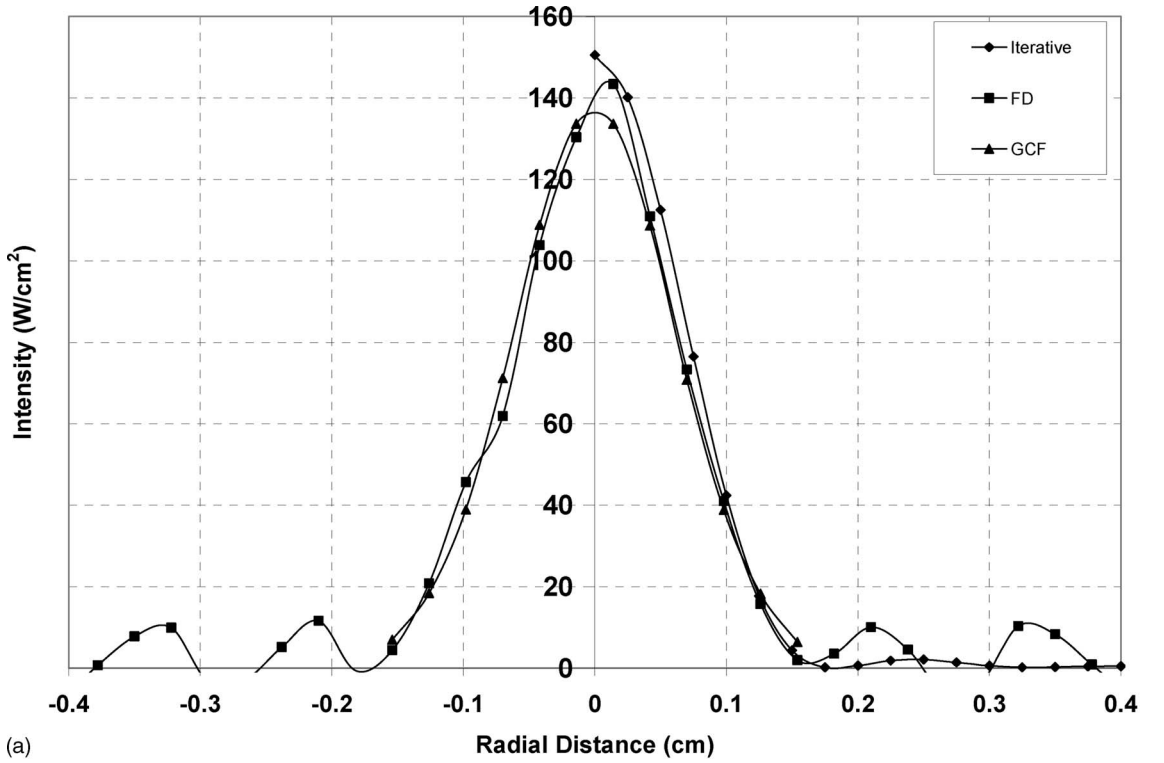
	Transducers		
Parameters	HIFU-1	HIFU-2	HIFU-3
Transducer radius	5 cm	3.8	3.2
Operating frequency	1.5 MHz	1.107 MHz	1.1 MHz
Focal distance	15 cm	11 cm	6.264 cm

Figure 4 pertains to the transducer HIFU 1, operated at a power of 5 W. The intensity [Fig. 4(a)] and normalized intensity [Fig. 4(b)] profiles in the focal plane are plotted for the GCF and FD methods, along with the iterative approach. The GCF intensity value on axis is less than that for the iterative method by about 10%, and the FD method is less than iterative by about 6%. The FD profile is not symmetric across the centerline, owing to the asymmetry of the streaming velocity in the illuminated sheet. Several sidelobes (at $r \approx 0.21$ mm, 0.33 mm) are present in the FD profile, but they do not match the location of the low magnitude lobe captured by the iterative approach at $r \approx 0.25$ mm. From the normalized profile [Fig. 4(b)], it can be seen that the quarter-power (−6 dB) beam width in the focal plane is approximately 2.1 mm, as predicted by all three streaming methods.

The axial dependence of the intensity for HIFU-1 is portrayed in Figs. 5(a) and 5(b). The amount of velocity data acquired (using our camera field of view setting) enabled direct intensity calculations only in the focal and postfocal regions. Both direct methods slightly underpredict the inverse-method intensity near the focus and overestimate it beyond about 8 mm from the focus. The axial −6 dB falloff distance postfocally is about 1.2 cm, according to all of the methods [Fig. 5(b)].

In Fig. 6, the transducer is HIFU-2, operated at a power of 3.6 W. The intensity [Fig. 6(a)] and normalized intensity [Fig. 6(b)] in the focal plane are plotted. The FD approach yields the slimmest radial profile. The GCF and iterative method profiles are similar to each other and wider than FD, but slimmer than the hydrophone measurements. (To reduce congestion, the symmetric iterative-method profile is shown only on the right half of the graph.) As with HIFU-1, FD shows a secondary maximum ($r \approx 0.17$ mm) that is not predicted by the other methods. The variation (maximum–minimum) in the prediction of on-axis intensity for the four methods is about 20%. The −6 dB beamwidths, derived from the relative-intensity profiles of Fig. 6(b), are about 3.2 mm (hydrophone), 2.7 mm (iterative), 2.6 mm (GCF), and 2.2 mm (FD).

When compared to hydrophone and iterative-method values, the axial dependence of the intensity is predicted reasonably well by the direct methods within a few mm of the focal region (Fig. 7). However, both direct methods manifest a much more rapid falloff with axial distance from the focal zone than the iterative or experimental methods. Insufficient data preclude computation of axial beamwidths, but it can be



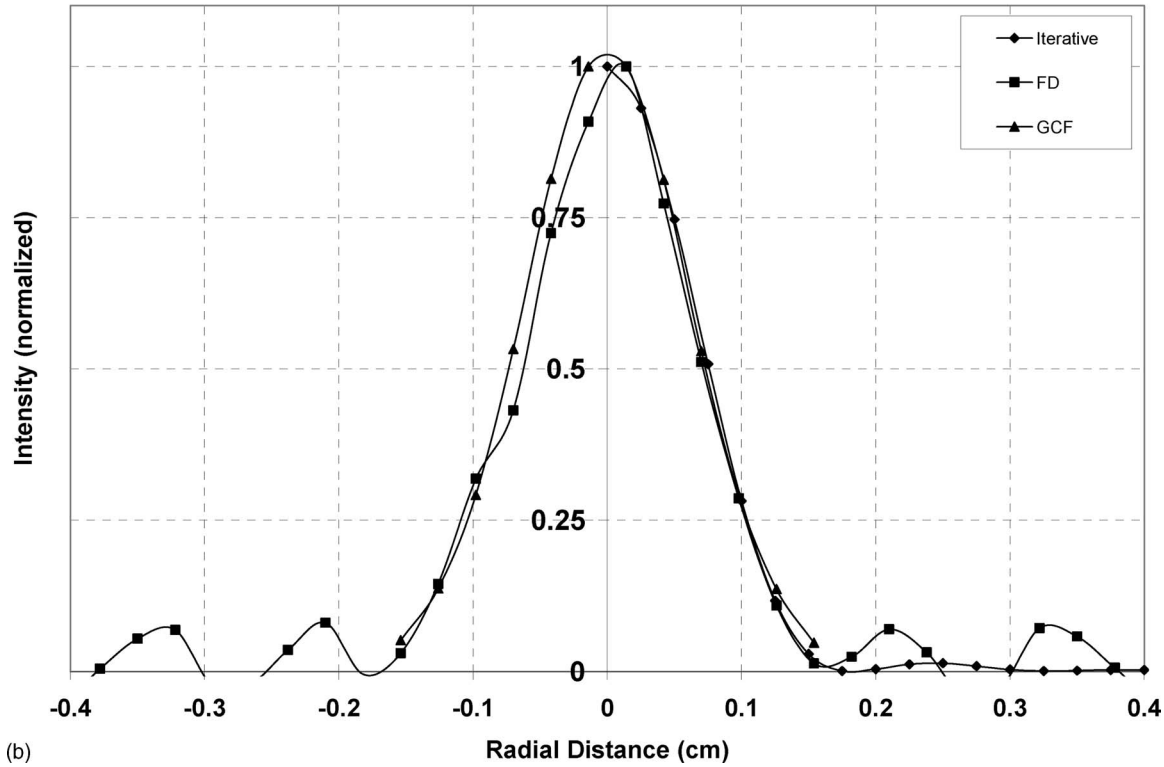
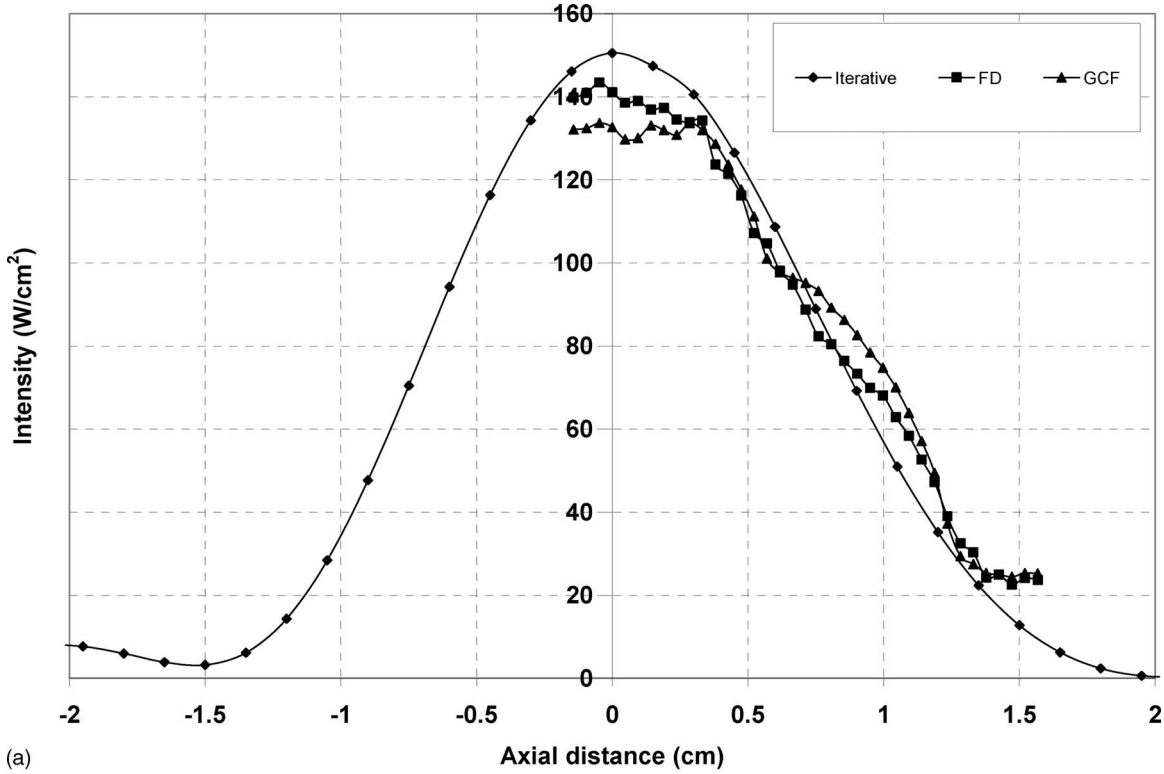


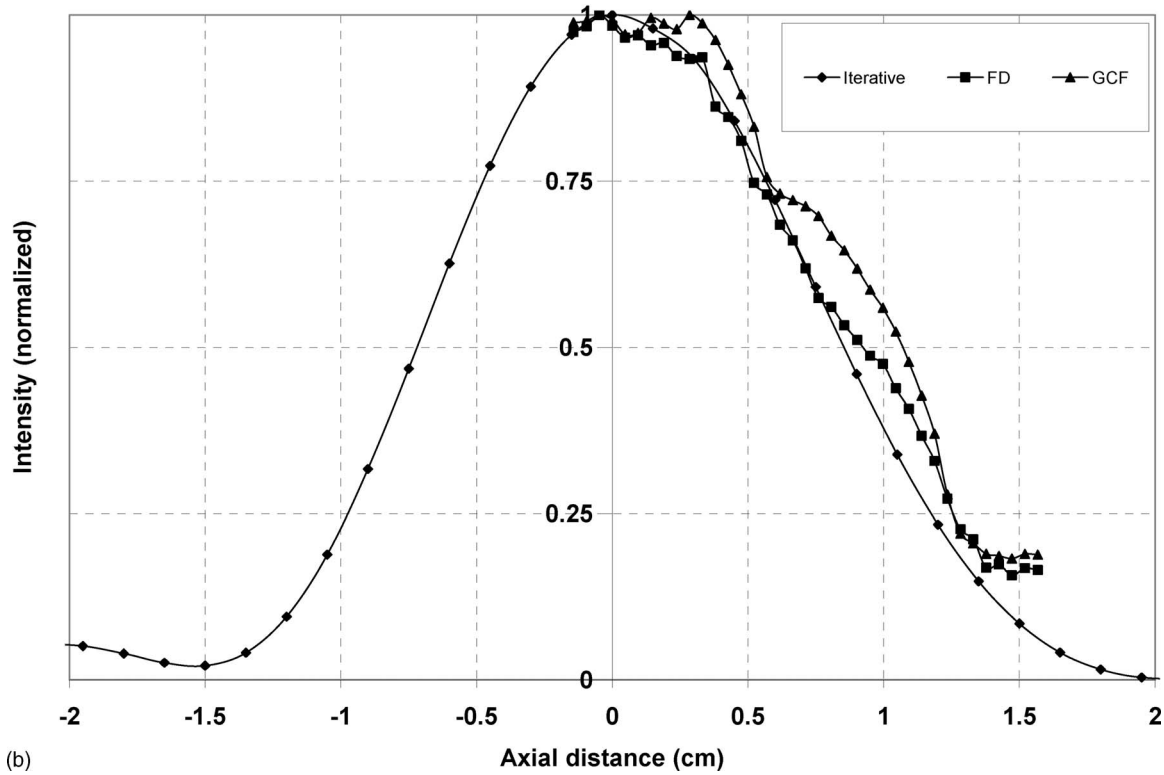
FIG. 4. (a) Acoustic intensity (W/cm^2) as a function of radial distance (cm) obtained from the finite-difference, Gaussian curve fit, and iterative methods. Acoustic power: 5 W; transducer: HIFU-1. (b) Normalized acoustic intensity as a function of radial distance (cm) obtained from the finite-difference, Gaussian curve fit, and iterative methods. Acoustic power: 5 W; transducer: HIFU-1.

said [Fig. 7(b)] that prefocally the axial intensity is half of its peak value about 8 mm in front of the peak for the hydrophone measurements, 7 mm for the iterative method, and about 6 mm for GCF.

For HIFU-3 operated at 5 W of power, all methods were quite close in their predictions of radial intensity profiles in the focal plane [Figs. 8(a) and 8(b)]. The hydrophone measurements were about 10% below the three streaming-based methods in the determination of on-axis intensity. FD predicted a noticeable secondary maximum once again, though this time the location better matched the smaller secondary maximum predicted by the iterative method. The −6 dB



(a)



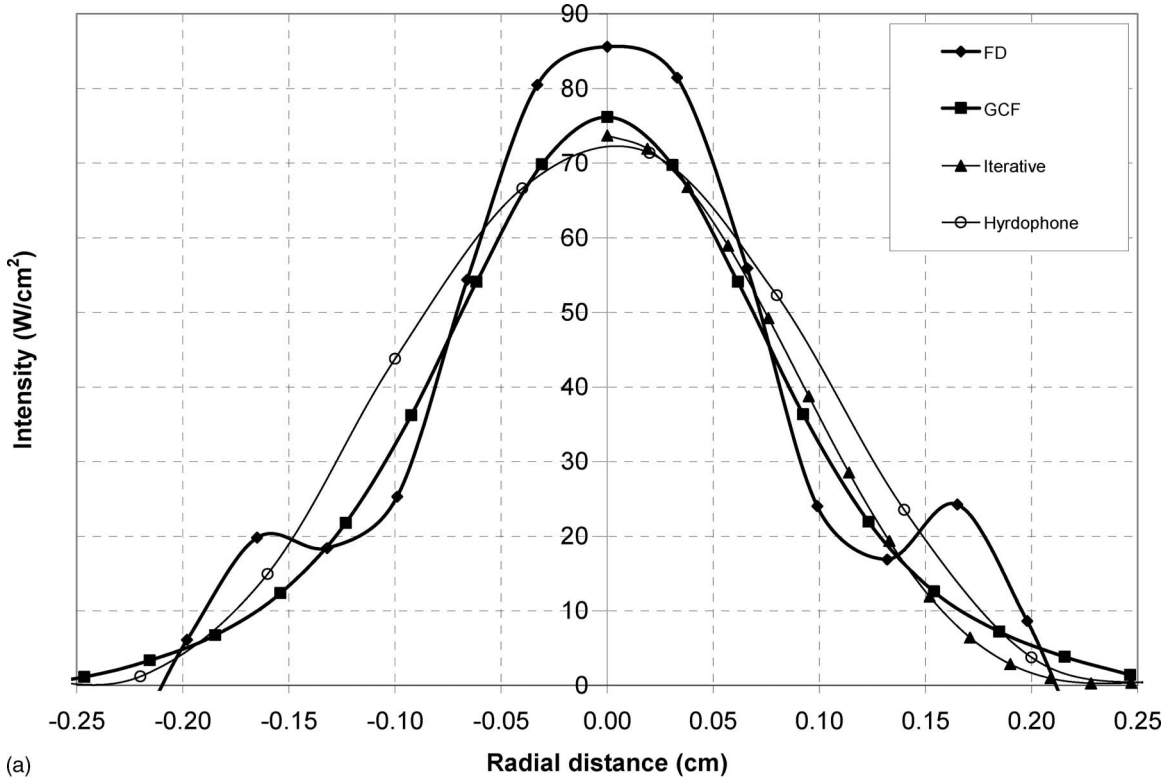
(b)

FIG. 5. (a) Acoustic intensity (W/cm^2) as a function of axial distance (cm) obtained from the finite-difference, Gaussian curve fit, and iterative methods. Acoustic power: 5 W; transducer: HIFU-1. (b) Normalized acoustic intensity as a function of axial distance (cm) obtained from the finite-difference, Gaussian curve fit, and iterative methods. Acoustic power: 5 W; transducer: HIFU-1.

beamwidths derived from Fig. 8(b) are 1.7 or 1.8 mm for the streaming-based methods, and about 2.1 mm for the hydrophone measurements.

The HIFU-3 axial intensity profiles derived from the various approaches also match closely [Figs. 9(a) and 9(b)]. The profile for the hydrophone measurements is slightly wider, with a −6 dB width of about 1.4 cm, compared with 1.2 or 1.3 cm for the streaming-based methods. (Some extrapolation on the prefocal side is required to estimate −6 dB intensity for FD and GCF.)

The inherent jaggedness of the direct-method plots can be reduced by fitting the intensity values to a smooth curve at



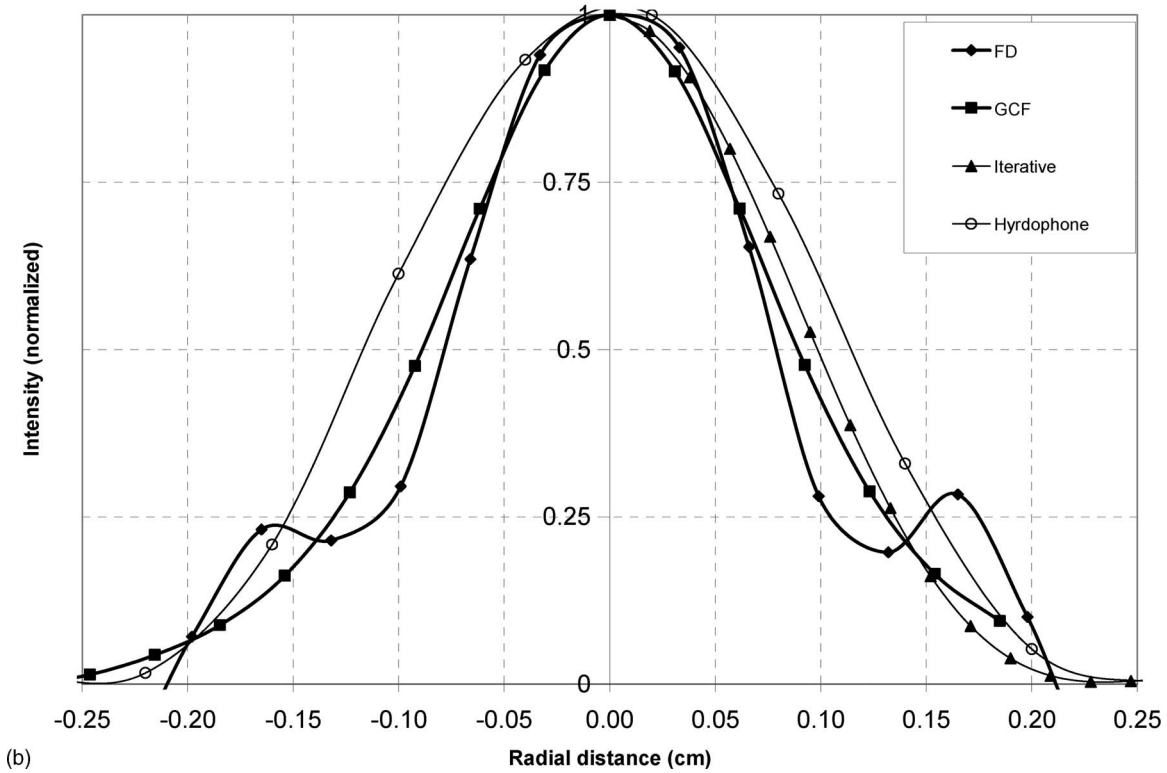
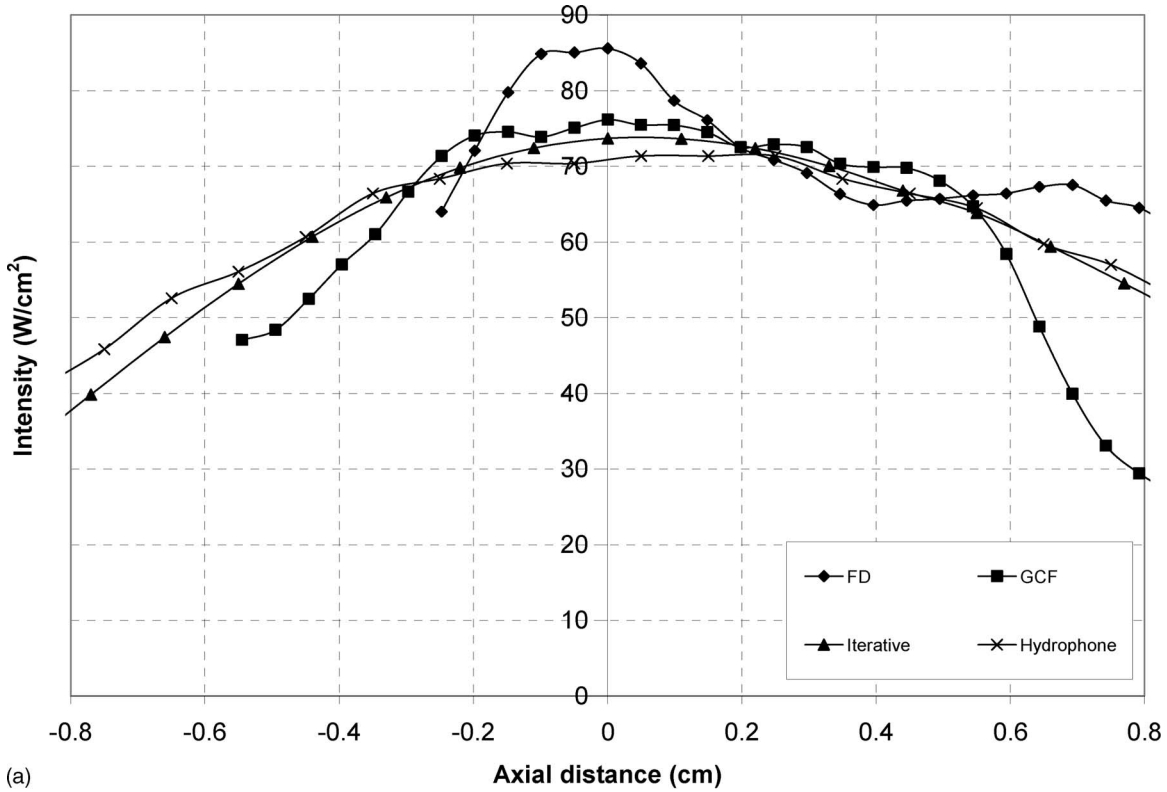


FIG. 6. (a) Acoustic intensity (W/cm^2) as a function of radial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurement. Acoustic power: 3.6 W; transducer: HIFU-2. (b) Normalized acoustic intensity as a function of radial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurement. Acoustic power: 3.6 W; transducer: HIFU-2.

the termination of the FD or GCF procedures. In Fig. 10, the GCF results from Fig. 9 have been fitted with a fifth-order polynomial. Also shown are the GCF results from Fig. 9 and the hydrophone measurements. Relative to the hydrophone measurements, the fitting polynomial does not produce a substantially more accurate representation. However, the fitting polynomial accurately extends the GCF results prefocally, in the sense of closely following the values derived from the iterative technique. [Iterative values shown in Fig. 9(a)].



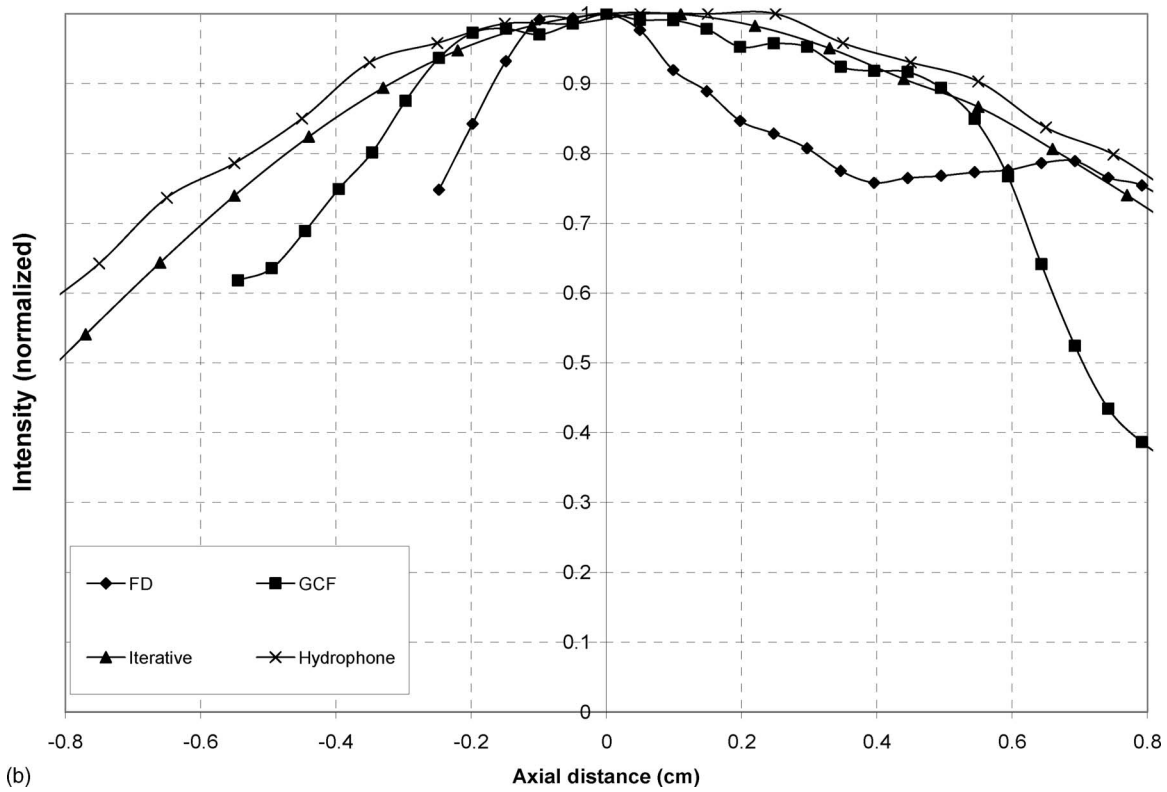
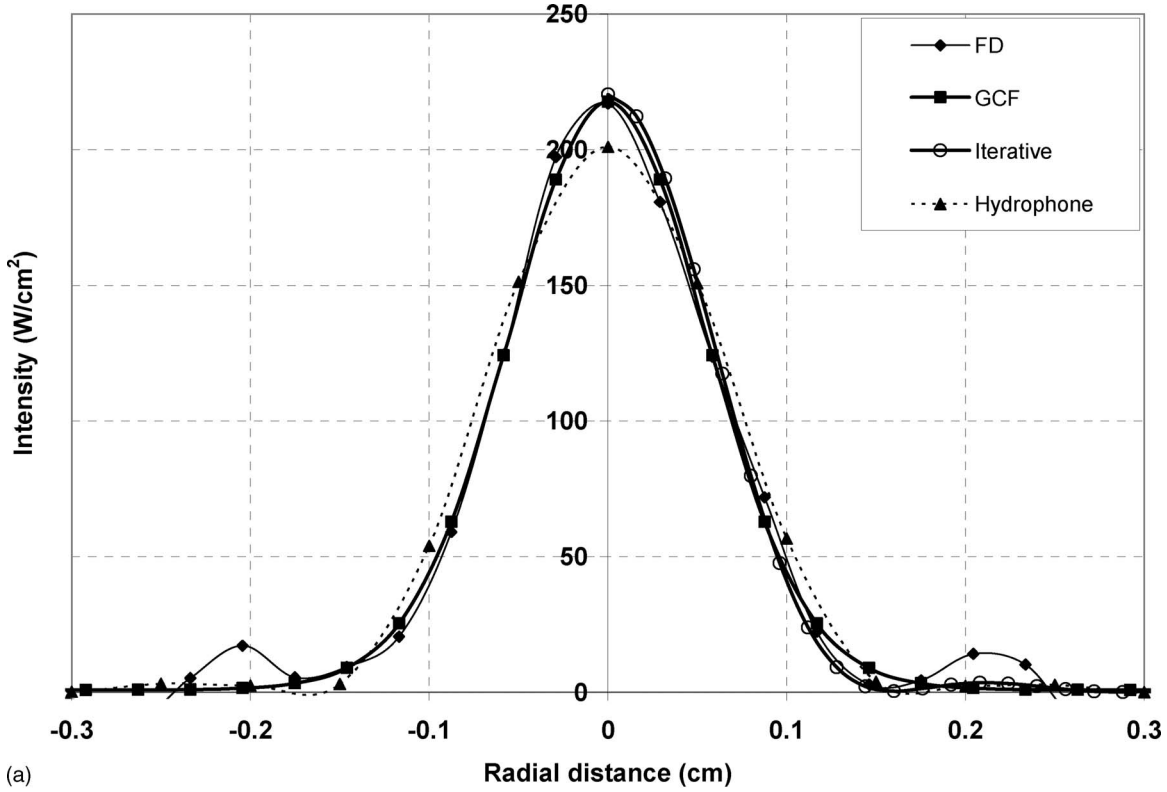
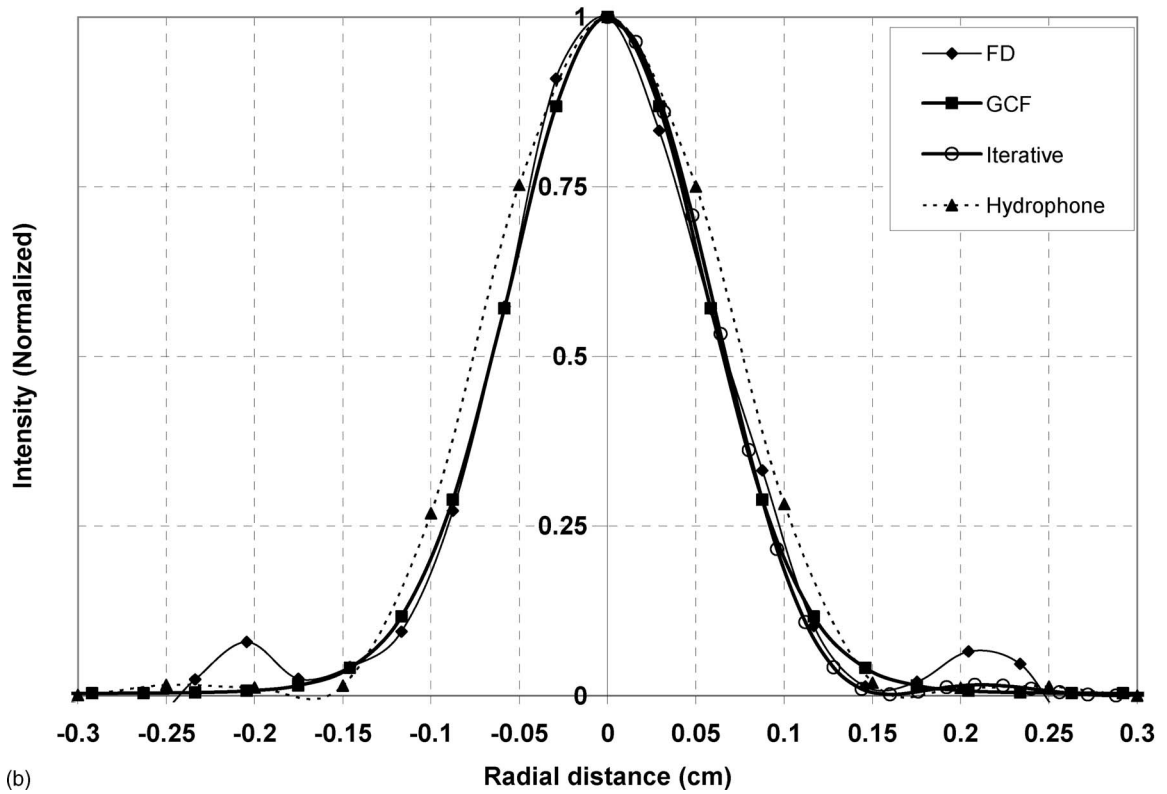


FIG. 7. (a) Acoustic intensity (W/cm^2) as a function of axial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurement. Acoustic power: 3.6 W; transducer: HIFU-2. (b) Normalized acoustic intensity as a function of axial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurement. Acoustic power: 3.6 W; transducer: HIFU-2.

The intensity computations thus far have involved transducer powers below 5 W. Streaming measurements were also made at transducer powers between 5 and 30 W. In Fig. 11, radial intensity profiles in the focal plane are presented for HIFU-3 operated at 30 W of acoustic power. The direct methods yield an on-axis intensity of approximately 1100 W/cm^2, compared with about 1300 W/cm^2 for the iterative method. Both the direct methods and the iterative



(a)

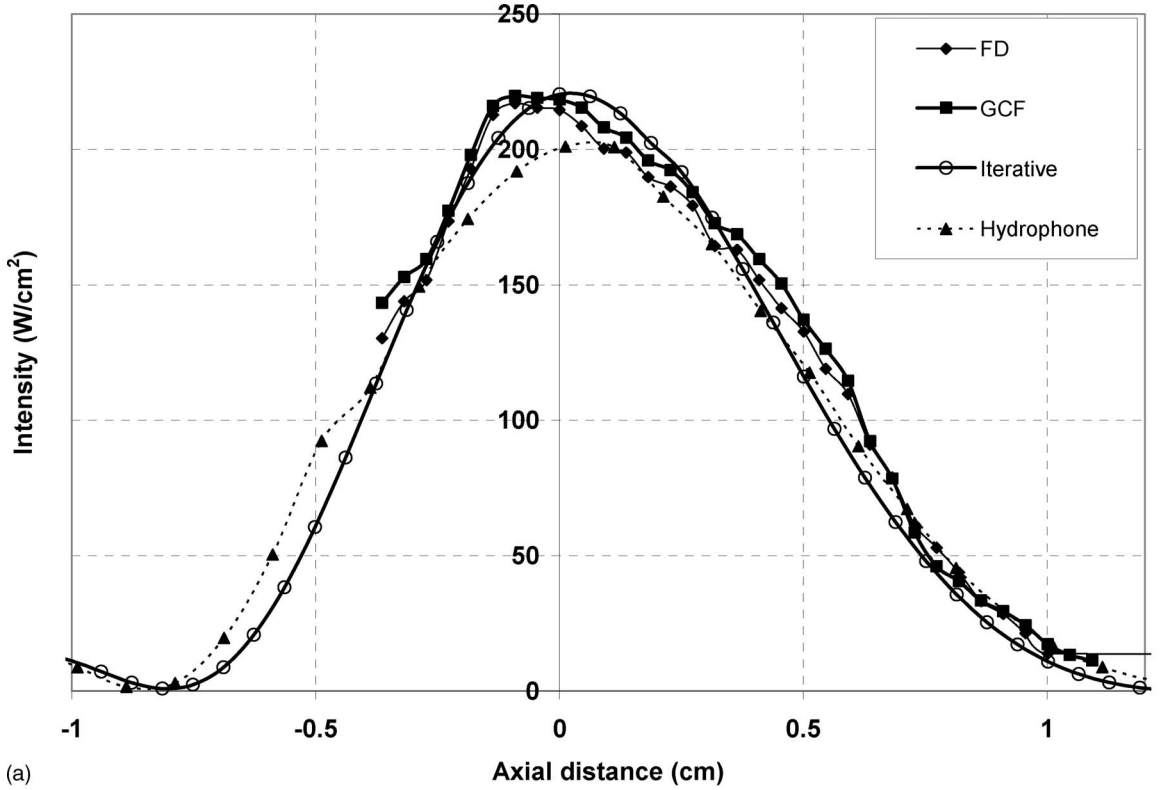


(b)

FIG. 8. (a) Acoustic intensity (W/cm^2) as a function of radial distance (cm) obtained from the methods: (i) finite difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurements. Acoustic power: 5 W; transducer: HIFU-3. (b) Normalized acoustic intensity as a function of radial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurements. Acoustic power: 5 W; transducer: HIFU-3.

approach produce −6 dB beamwidths of 1.9 mm ± 0.1 mm. The FD and GCF profiles match particularly well if the asymmetric FD profile is shifted to the right about 1 mm.

To check the consistency of the direct methods as the streaming speed—and hence the hydrodynamic nonlinearity—was increased, the intensity was computed as



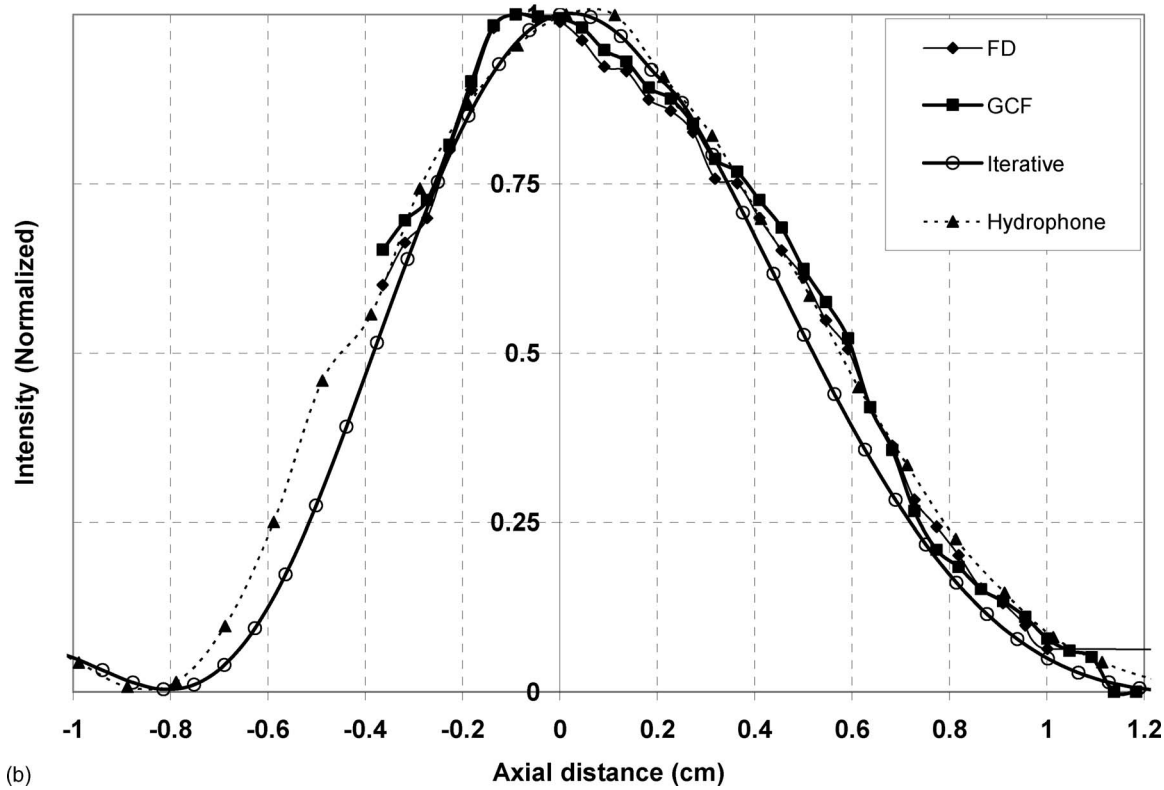


FIG. 9. (a) Acoustic intensity (W/cm^2) as a function of axial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (iii) iterative, and (iv) hydrophone measurements. Acoustic power: 5 W; transducer: HIFU-3. (b) Normalized acoustic intensity as a function of axial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, (ii) iterative, and (iv) hydrophone measurements. Acoustic power: 5 W; transducer: HIFU-3.

a function of acoustic power. As Fig. 12 illustrates, a linear dependence of peak intensity upon acoustic power was observed over the range of powers used in our experiments.

IV. DISCUSSION

Peak intensities and beam widths derived from the FD and GCF direct methods generally matched those derived

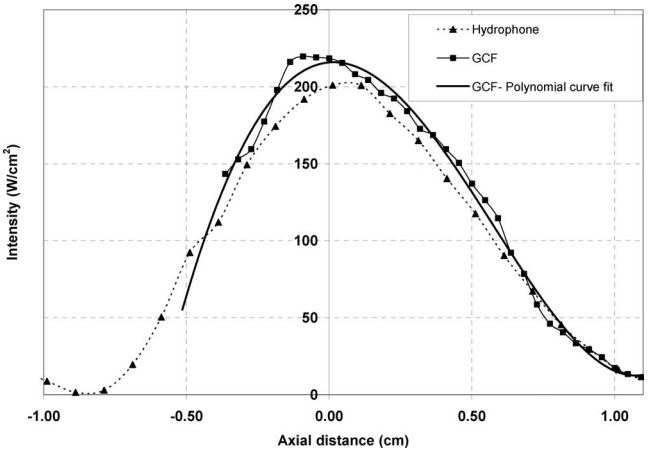


FIG. 10. Radial intensity profiles obtained from (i) GCF method, (ii) hydrophone measurements, and (iii) polynomial curve fit of GCF results. Acoustic power: 5 W; transducer: HIFU 3.

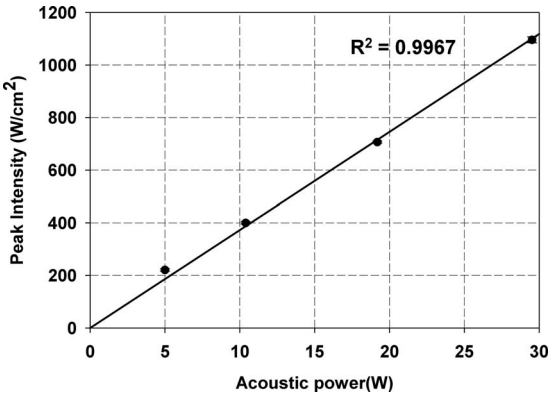


FIG. 12. Peak acoustic intensity (W/cm^2) as a function of input acoustic power (W) for HIFU 3 transducer.

from the iterative method and hydrophone measurements to within 20%. Even for the larger discrepancies with experiment [e.g., Fig. 6(a)], it should be kept in mind that the intensities derived from hydrophone measurements can possess uncertainties in excess of 20% (Hariharan *et al.*, 2008). Given the comparable accuracy of the direct methods to other characterization techniques, their noninvasive nature, and the fact that the direct methods require only seconds of computation time, the direct methods applied to acoustic streaming data provide an attractive option for transducer characterization.

The terms incorporated into the Navier–Stokes operator of Eq. (2.6) include the convective derivative $\overline{u_z}\ \overline{\partial u_z/\partial z}$. In a successive-approximation treatment of the streaming flow (Nyborg, 1998), the streaming velocity is formally second order (the acoustic quantities of frequency ω being first order), and this convective term is therefore fourth order. In successive-approximation treatments this term is sometimes neglected in the governing equations (Nyborg, 1998). However, when the governing Eq. (2.6) is nondimensionalized using the width d_s of the streaming jet and a typical velocity U_s of the jet, the convective derivative is proportional to streaming Reynolds number $U_s d_s/\nu$ (ν being the kinematic viscosity of the streaming fluid). Depending upon the viscosity of the streaming medium, the Reynolds number for the streaming jet can be on the order of 100 for ultrasound beams in the moderate-intensity regime. The convective derivative can therefore be significant despite being formally higher order, and, as noted by Lighthill (1978), should be retained in streaming calculations except at low power. In the case of the focused transducers considered in this paper, and additional factor is present. Near the geometric focus the intensity is maximum and the axial derivative of the intensity is zero. Likewise, the axial derivative of the velocity u_z decreases to zero at a location near the focus, thus diminishing the importance of the convective term. In terms of numerical magnitudes, we found the convective term $\rho_0\ \overline{u_z}\ \overline{\partial u_z/\partial z}$ to be on the order of 10% as large as the viscous stresses in the Navier–Stokes equations, in the region of interest centered around the focus. We feel it is advisable to retain the convective term in the moderate-intensity regime, and higher power levels. Below the moderate-intensity regime, the convective term could likely be neglected near the focal region.

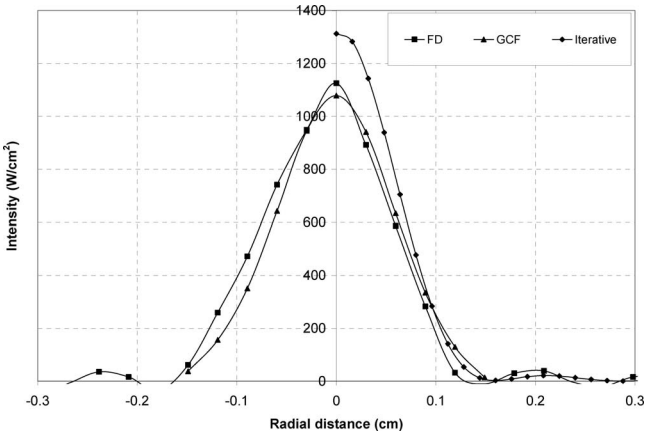


FIG. 11. Acoustic intensity (W/cm^2) as a function of radial distance (cm) obtained from the methods: (i) finite-difference, (ii) Gaussian curve fit, and (iii) iterative. Acoustic power: 30 W; transducer: HIFU 3.

A limitation of the direct methods can be seen in the axial-intensity characterization of HIFU-2 [Fig. 5(a)]. Because DPIV data were available over an axial range of only about 2 cm, the prefocal intensity distribution could not be characterized. Generally speaking, a disadvantage of the direct techniques are the requirement of accurate velocity data in the region of interest. With the iterative method, streaming velocity data from the focal region are essentially used to "calibrate" the acoustic propagation code, i.e., select the set of operational parameters (transducer power, focal length, steering angle; absorption of medium,…) that best predicts the streaming motion. Once calibrated, the propagation code can predict the intensity at any location of interest, even though streaming data may be inaccurate or unavailable at that location. With the direct methods, inaccuracies in the velocity field can yield large inaccuracies in the intensity predictions. If intensity predictions outside of the camera field of view are desired using a direct method, DPIV experiments can be repeated with the camera view centered on the desired location, in order to enhance the accuracy of the streaming measurements.

The optimal step size for finite difference calculations [Eqs. (2.9) and (2.10)] correlates well with the size of the

interrogation window used to process the PIV images. In the PIV technique, images obtained from the charge coupled device camera are usually divided into small subsections called interrogation areas whose size determines the resolution of the streaming technique. The size of the interrogation area depends on factors such as seeding density, particle size, and the image magnification factor. The requirement for the interrogation spot size is that it should be large enough to contain at least 5–10 particle images with an ideal particle diameter of about 2 pixels (Prasad, 2000). For this study, the optimal size of the interrogation area is estimated to be around ~0.03 cm, the same as the optimal grid size.

Using the optimal grid size, the FD method was able to compute reasonably accurate velocity gradients and intensities within the focal region. However, outside the focal region, oscillations can be noticed in the intensity profiles (Figs. 4, 6, and 8). These oscillations can be viewed as a consequence of insufficiently accurate velocity data, due to the presence of steep velocity gradients in the streaming field. In PIV, velocity is measured by tracking the particle patterns captured from two successive images which are separated by a preset time interval. Ideally, for accurate measurement of velocity, the particle displacement during the time interval should be around ~8 pixels. Usually, the time interval is selected based on the displacement of the fastest moving particle group. In this case, it was selected based on the displacement of particles in the focal region. Consequently, particles in the much lower velocity region outside the focus were not displaced far enough for the cross-correlation algorithm to calculate the velocity accurately. As noted above, one way to acquire velocity data sufficiently accurate to resolve sidelobes would be to refocus the camera at the sidelobe location and acquire an additional set of data.

A second way to overcome problems imposed by high velocity gradients is to increase the viscosity of the medium. The higher viscosity serves to radially diffuse momentum, thereby reducing the radial velocity gradient. The increased viscosity also reduces the strength of the hydrodynamic nonlinearity in the streaming jet, making it more stable. On the other hand, increasing the viscosity can also increase the acoustic absorption, which enhances the driving force and the streaming velocity. The net effect of increasing the viscosity, therefore, can be difficult to predict (Hariharan *et al.*, 2008).

While the FD method can in principal resolve intensity sidelobes, the same is not true for the GCF technique. The Gaussian-based functions (2.7) and (2.8) are meant to represent the field only in the focal region. Because no oscillatory behavior is built into the Gaussian functions, they cannot predict the location or magnitude of any sidelobes of the intensity distribution.

The Gaussian function in Eq. (2.7) models the second derivative of the axial velocity. However, near $r=0$, the other radial term in the Laplacian, $(1/r)(\partial u_z/\partial r)$, is also well represented by Eq. (2.7). [Since $\partial u_z/\partial r=0$ at $r=0$, the first term in the Taylor series representation for both $(1/r)(\partial u_z/\partial r)$ and $\partial^2 u_z/\partial r^2$ is the same, $(\partial^2 u_z(r=0)/\partial r^2)$.] Additionally, the axial derivatives in Eq. (2.8) are generally small compared with those in the radial direction. It can be approximately stated, then, that the Gaussian profile models the radial intensity distribution as well as the second derivative.

While the Gaussian-based profile Eq. (2.8) accurately fits the streaming-velocity distribution (Fig. 2), other fitting functions can be devised as well. Polynomial functions—including quadratic functions—were found to accurately represent the data of Fig. 2. However, these polynomial functions sometimes yielded very inaccurate intensity predictions. The quadratic function, for example, yields an intensity distribution that is a constant function of radial position. Hence, low mean-square error between the streaming-velocity data and any proposed fitting function cannot be the sole criterion for selection of a fitting function. We recommend that comparisons with the iterative method of Hariharan *et al.* (2008) be performed to insure that the derivatives defining the viscous stress are accurately determined.

The lower accuracy of the direct methods applied to HIFU-2 (Figs. 6 and 7) than the other transducers may be due to the low streaming velocities associated with HIFU-2. The transducer input voltage was the same for the three transducers, but the lower efficiency of HIFU-2 resulted in a lower acoustic power and a slower moving streaming fluid. In our DPIV system, velocities below about 1 cm/s were probably not resolved with sufficient accuracy to employ finite-differencing approaches. In practice, these limits of applicability of finite-differencing methods can be discerned through comparisons with the iterative method.

Curve fitting after application of the GCF or FD techniques (Fig. 10) did not seem to appreciably affect the accuracy of the methods, but is visually appealing. It also allows for simple interpolation and extrapolation of the results, e.g., for calculations of heating rates or radiation force.

The agreement between the direct methods and the iterative method for HIFU-3 at high power (Fig. 11) is not as good as it is at low power (Fig. 8). At the intensity values of Fig. 11, which exceed 1000 W/cm^2, nonlinear acoustic propagation effects are likely emerging. We note that both the iterative method and the direct methods contain a source term [proportional to F_z defined in Eq. (2.3)] that cannot adequately treat significant acoustic nonlinearities; the source should actually be a sum that includes the additional modes generated by nonlinear propagation. The source term should additionally reflect the frequency dependence of the absorption. The effects of acoustic nonlinearities on the direct and iterative approaches is not known and requires further study. Still, the reasonable agreement between the methods featured in Fig. 11 indicates that the direct method remains accurate as the velocity (and Reynolds number) of the streaming jet increases. In other words, while the importance of the hydrodynamic nonlinearity increases, the direct techniques are able to adequately capture the transfer of momentum from the acoustic field to the hydrodynamic field. This feature is also confirmed by the linearity of the results in Fig. 12: Provided the acoustic propagation can be assumed linear (and locally plane wave), we would expect the peak intensity to be proportional to the transducer acoustic power, regardless of the speed of the streaming fluid.

The direct techniques have been presented in the context of axisymmetric transducers. For three-dimensional intensity

fields, the FD method can be applied with little modification, as long as three-dimensional velocity information with sufficient accuracy is available. The operators in Eq. (2.6) need only be written in three-dimensional form, and the finite differencing performed in three coordinate directions. For the GCF method, the Gaussian fitting function might need to be replaced with a function more appropriate for the transducer geometry of interest. With either direct method, little increase in computation time would result. For the iterative method, on the other hand, a substantial increase in computation time would be incurred in moving from axisymmetric to three-dimensional geometries.

V. CONCLUSION

The two direct methods presented in this paper represent additional tools for noninvasively characterizing HIFU transducers. They are much faster than the previously reported streaming technique that utilizes an optimization process. They are also simpler and potentially cheaper, in the sense that no software is required for the solution of the acoustic propagation equations or the equations of fluid flow. A limitation of the direct techniques is that they provide intensity estimates only where accurate streaming velocity is available. They also require some care to limit the instabilities associated with numerical differentiation of experimental data. For example, any changes to the finite differencing interval (FD method) or the fitting function (GCF) can significantly increase the error associated with the direct technique. Any changes to the algorithm parameters should therefore be validated through comparison with the iterative technique or experimental measurements.

ACKNOWLEDGMENTS

Assistance from Ronald Robinson on the DPIV experiments is gratefully acknowledged. We also thank Subha Maruvada for performing the radiation force balance experiments and Jerry Harris for assisting with the hydrophone measurements.

Abramowitz, M., and Stegun, I. A. (**1972**). *Handbook of Mathematical Functions* (Dover, New York).

Gerald, C. F. (**1978**). *Applied Numerical Analysis* (Addison-Wesley, Reading, MA).

Hariharan, P., Myers, M. R., Robinson, R. A., Maruvada, S. H., Sliwa, J., and Banerjee, R. K. (**2008**). "Characterization of high intensity focused ultrasound transducers using acoustic streaming," J. Acoust. Soc. Am. **123**, 1706–1719.

Lighthill, J. (**1978**). "Acoustic streaming," J. Sound Vib. **61**, 391–418.

Mathworks, Inc. (**2002**). *Matlab Manual* (Mathworks, Inc., Natick, MA), ver. 6.5

Nyborg, W. L. (**1965**). *Physical Acoustics* (Academic, New York), Vol. **2(B)**.

Nyborg, W. L. (**1998**). "Acoustic streaming," in *Nonlinear Acoustics*, edited by Hamilton and Blackstock (Academic, San Diego).

Prasad, A. K. (**2000**). "Particle image velocimetry," Curr. Sci. **79**(1), 51–60.

Shaw, A., and ter Haar, G. R. (**2006**). "Requirements for measurement standards in High Intensity Focused (HIFU) Ultrasound Fields," National Physical Laboratory (NPL) Report No. ISSN 1744-0599.

Wu, J., and Du, G. (**1993**). "Acoustic streaming generated by a focused Gaussian beam and finite amplitude tonebursts," Ultrasound Med. Biol. **19**, 167–176.

Development and characterization of a blood mimicking fluid for high intensity focused ultrasound

Yunbo Liu and Subha Maruvada
Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland 20993

Randy L. King
Department of Bioengineering, Stanford University, Stanford, California 94305

Bruce A. Herman and Keith A. Wear
Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland 20993

(Received 4 February 2008; revised 9 June 2008; accepted 10 June 2008)

A blood mimicking fluid (BMF) has been developed for the acoustic and thermal characterizations of high intensity focused ultrasound (HIFU) ablation devices. The BMF is based on a degassed and de-ionized water solution dispersed with low density polyethylene microspheres, nylon particles, gellan gum, and glycerol. A broad range of physical parameters, including attenuation coefficient, speed of sound, viscosity, thermal conductivity, and diffusivity, were characterized as a function of temperature (20–70 °C). The nonlinear parameter B/A and backscatter coefficient were also measured at room temperature. Importantly, the attenuation coefficient is linearly proportional to the frequency (2–8 MHz) with a slope of about 0.2 dB cm^{-1} MHz^{-1} in the 20–70 °C range as in the case of human blood. Furthermore, sound speed and bloodlike backscattering indicate the usefulness of the BMF for ultrasound flow imaging and ultrasound-guided HIFU applications. Most of the other temperature-dependent physical parameters are also close to the reported values in human blood. These properties make it a unique HIFU research tool for developing standardized exposimetry techniques, validating numerical models, and determining the safety and efficacy of HIFU ablation devices. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2956469]

PACS number(s): 43.80.Ev, 43.80.Cs, 43.80.Vj [CCC] Pages: 1803–1810

I. INTRODUCTION

In recent years, high intensity focused ultrasound (HIFU) has re-emerged as a minimally invasive and nonionizing therapeutic modality for the treatment of a variety of solid tumors (both malignant and benign), including liver (Kennedy *et al.*, 2004, Jolesz *et al.*, 2004), prostate (Chapelon *et al.*, 1999; Chaussy and Thuroff, 2003), breast (Hynynen *et al.*, 2001), and uterine fibroids (McDannold *et al.*, 2006). Pioneering mechanism studies have demonstrated that HIFU, with spatial-peak temporal-average intensity (I_{SPTA}) between 500 and 10 000 W cm^{-2} at the beam focus, can produce necrotic lesions in deep-sited tissue through thermal ablation ($T>55$ °C, $t<30$ s) and inertial cavitation with minimal damage to the surrounding tissues (Bailey *et al.*, 2003; ter Haar and Coussios, 2007). New biomedical applications of this therapeutic method are also being developed, such as cauterization of deep bleeding vessels for noninvasive hemorrhage control (Vaezy and Zderic, 2007) or arterial thermal occlusion in perfused tissue (Ichihara *et al.*, 2007). During all these HIFU therapy applications, the perfusion of blood flow in the targeted area can affect the required acoustic power, thermal deposition pattern, and corresponding lesion formation process at the focus (Bailey *et al.*, 2003; Vaezy and Zderic, 2007). To better understand the HIFU tissue heating process with convective vascular structures, *in vitro* (Huang *et al.*, 2004), *in vivo* (Chen *et al.*, 1993), and clinical studies (Wiart *et al.*, 2007) have been conducted under various blood flow conditions. Meanwhile, extensive numerical simulations have indicated a reduced peak temperature near or inside the blood vessel due to the perfusion cooling and acoustic streaming (Bailey *et al.*, 2003; Huang *et al.*, 2004, Hariharan *et al.*, 2007). Therefore, blood flow in major vessels may play an important role for effective HIFU monitoring and treatment, especially during ultrasound Doppler imaging-guided therapy (Martin *et al.*, 1999).

Despite all these encouraging achievements, measurement and regulation guidelines are still not standardized to determine the safety and effectiveness of such therapeutic devices. In order to develop measurement techniques, validate theoretical models, and characterize specific HIFU ablation devices, a HIFU blood mimicking fluid (BMF) with well characterized physical properties needs to be developed. Relevant engineering testing can then be conducted before larger scale animal or clinical studies are performed. The desired BMF should have acoustic propagation and scattering characteristics similar to human blood, both for the "linear" range and the high intensity "nonlinear" regime of ultrasound sonication. Since temperature rise is the principal therapeutic mode for this technology, the BMF developed must also have similar thermal properties to human tissues. Of particular interest is that the BMF should be capable of functioning stably under the severe HIFU exposures, such as

high temperature, strong radiation force, and acoustical streaming, without significant property damage.

In the past two decades, several BMFs, mainly water-based solutions dissolved with various scattering particles, have been reported (Boote and Zagzebski, 1988; Oates, 1991; Ramnarine *et al.*, 1998; Samavat and Evans, 2006). However, these BMFs were designed primarily for ultrasound Doppler imaging studies emphasizing proper backscattering at room temperature. The attenuation coefficients of these BMFs (0.05 dB cm^{-1} MHz^{-1}) are much lower than that of the real blood (0.2 dB cm^{-1} MHz^{-1}) and the other temperature-dependent physical properties are not well characterized (Ramnarine *et al.*, 1998; Huang *et al.*, 2004; Samavat and Evans, 2006). Therefore, they may not be appropriate for conducting measurement and simulation studies during HIFU device evaluation due to the large temperature rise, nonlinear harmonic absorption, and thermal perfusion by large vessels. Most recently, animal blood samples have also been used in a pulsatile HIFU flow phantom system (Greaby *et al.*, 2007). The biohazard risk of biological media, short shelf life, as well as potential blood cell denature-associated property variation can complicate the device calibration results and hinder the direct comparison between the measurement and the numerical simulation.

The BMF devised in this work is a degassed water solution mainly dispersed with low density polyethylene (LDPE) and nylon microspheres. This BMF was investigated experimentally as a potential calibration tool for HIFU ablation devices. The temperature dependence of acoustic attenuation, speed of sound, thermal diffusivity, thermal conductivity, and viscosity were systematically studied from 20 to 70 °C based on the established measurement methods. As another critical aspect for HIFU applications, the nonlinear parameter B/A and the ultrasound backscatter coefficient of the BMF were also characterized at room temperature. These investigations demonstrate that the current BMF can be manufactured conveniently with acoustical and thermal properties close to human blood and, of particular importance, could be used repeatedly under HIFU temperature conditions ($T>55$ °C).

II. MATERIALS AND METHODS

A. Blood mimicking fluid production

In order to simulate the ultrasound attenuation of human blood, LDPE microspheres ($D=6$ μm, CL-1080, Sumitomo Seika, Japan) and nylon particles ($D=8-12$ μm, SP-10, Kobo Product Inc, South Plainfield, NJ) are used as absorptive and scattering elements, respectively. LDPE has a reported linear frequency-dependent attenuation coefficient of 4 dB cm^{-1} MHz^{-1} in the 1–10 MHz range and close acoustical impedance (1.7×10^6 Pa s m^{-1}) to that of biological tissues (Preston *et al.*, 1991; O'Neill *et al.*, 1994). These advantages of LDPE make it a unique and attractive material for providing ultrasound absorption during HIFU ablation. Nylon particles have much higher acoustical impedance ($\sim3\times10^6$ Pa s m^{-1}) and have been used to simulate the scattering property of blood in Doppler ultrasound phantoms (Boote and Zagzebski, 1988). Both the LDPE and nylon microspheres can be dispersed uniformly in the solution by adding a nonionic, low-foam surfactant (Tergitol L64, Dow Chemical Co., Midland, Michigan). Glycerol (G7757, Sigma Co., St. Louis, MO) and gellan gum (Kelcogel CG-LA, CP Kelco, Atlanta, GA) are added to optimize the density and match the speed of sound, B/A, and viscosity with that of human blood. Density is critical for maximizing the LDPE and nylon particle buoyancy in the phantom solution. Gellan gum is an agarlike, nontoxic, polysaccharide powder produced from the bacterium *Pseudomonas elodea*. Gellan gum contains very low organic compound impurities so consequent bacterial contamination is less likely to happen. All the ingredients of the BMF have high temperature stability (>95 °C), which is critical for HIFU applications.

TABLE I. Blood mimicking fluid recipe.

Materials	Function	Content
Degassed H_2O	Substrate	100 ml
LDPE	Absorption	6 g (6 μm)
Nylon	Scattering	0.5 g (8–12 μm)
Gellan gum	Viscosity	0.1 g
Glycerin	Viscosity/density/(B/A)	10 ml
Low-foam surfactant	LDPE/nylon wetting	1 ml

Table I summarizes the specific additives and their corresponding function in the BMF recipe. A representative protocol for producing 100 ml of BMF is also described in the Appendix. No preservative additives were used in the current recipe since fresh BMF can be made easily for experiments. The BMF holders were two custom-designed cylindrical test chambers (inside diameter=4 cm; thickness=2 and 4 cm) with 25 μm LDPE membranes covering both ends (Fig. 1).

B. Measurement of physical properties

1. Ultrasound attenuation and speed of sound

For successful HIFU therapy, the focused acoustic wave should be largely absorbed within the target volume so that the local tissue temperature can be elevated rapidly due to the dissipative and viscous nature of biological tissues

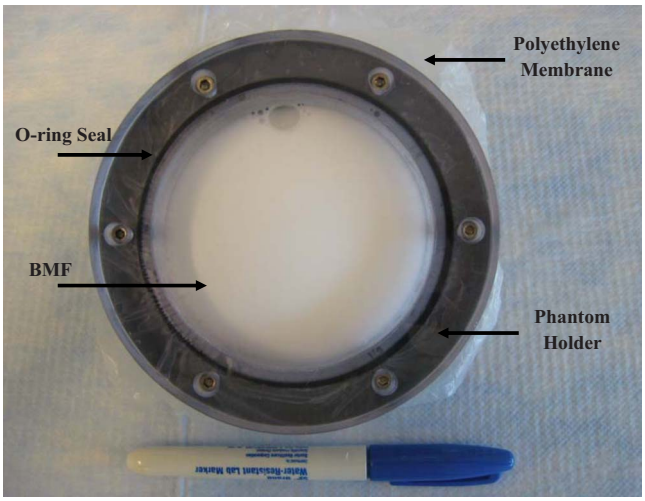


FIG. 1. (Color online) BMF in a 2 cm thick sample holder.

(Bailey *et al.*, 2003). The ultrasound attenuation coefficient (α, dB cm^{-1}), consisting of both scattering and absorption effects, is the principal physical parameter that determines the dissipation efficiency of acoustical energy within a specific medium. The attenuation coefficient has an approximately linear ultrasonic frequency dependence in normal human blood and can be expressed mathematically as $\alpha = a \cdot f^b$, where a is typically from 0.16 to 0.24 and b is from 1.19 to 1.23 (Duck, 1990). The speed of sound in human blood is around 1540–1590 $m \cdot s^{-1}$ at body temperature (Duck, 1990; Oates, 1991; Ramnarine *et al.*, 1998).

Through-transmission substitution techniques have been employed to quantify the frequency specific ultrasound attenuation and sound speed by several groups (Madsen *et al.*, 1999). In the current study, frequency- (2–8 MHz) and temperature-dependent (20–70 °C) ultrasound attenuation and sound speed of the HIFU BMF were investigated using an ultrasonic time delay spectrometry (TDS) system in a temperature-regulated water bath (Gammell *et al.*, 2007). The TDS measurement is a broadband, linear swept-frequency method that produces direct spectral information with immunity to multipath interference effects and enhanced signal-to-noise ratio due to its inherently narrowband signal acquisition. The attenuation (insertion loss) was measured by subtracting the spectrum (in decibel) transmitted through the 2 and 4 cm thick BMF samples in order to eliminate the reflection losses at the anterior and posterior interfaces. Also, the speed of sound in the BMF was obtained during the TDS measurement by considering the reference speed of sound in water as a function of temperature from 20 to 70 °C (Marczak, 1997).

2. Nonlinear parameter B/A

As a fundamental estimation of medium nonlinearity for finite amplitude wave propagation, the nonlinear parameter B/A plays an important role in enhancing lesion inception and growth during HIFU thermal ablation (Filonenko and Khokhlova, 2001). Several methods have been utilized to measure B/A in biological media (Law *et al.*, 1984; Dong *et al.*, 1999). In this experiment, a recently developed method based on the finite amplitude insert-substitution technique (Dong *et al.*, 1999) was employed to measure the second harmonic amplitudes following transmission of an unfocused tone burst having a fundamental frequency of 2.25 MHz through degassed water with and without a 2 cm BMF sample (Harris *et al.*, 2007). The ratio of these second harmonic amplitudes yields the B/A of the samples assuming the known B/A values for water (Duck, 1990). An advantage of this technique is that neither absolute pressure amplitude measurements nor diffraction corrections are required.

3. Backscatter coefficients and Doppler images

Accurate targeting and monitoring of blood flow and bleeding are critical for the development of HIFU hemostasis. Therefore, a key acoustical scattering property of the BMF was also evaluated. The ultrasound backscatter coefficient (cm^{-1} Sr^{-1}) was measured based on a broadband pulse-echo reference phantom technique (Wear, 1999). Basically, a 3.5 MHz, circular focused broadband transducer (Panametrics Inc., Waltham, WA) was driven by a pulser/receiver (Model 5800, Panametrics Inc, Waltham, WA) in transmit/receive mode. The transducer was focused at the middle of the 2 cm BMF samples. Received rf ultrasound echoes were digitized (8 bit, 10 MHz) using a 400 MHz oscilloscope (Model 9310C, LeCroy Corporation, Chestnut Ridge, NY) and stored on computer via general purpose interface bus board. The frequency spectra of the rf signals from different positions within the samples, as well as from a reference phantom (Wear, 1999), were then averaged and processed off-line following an established protocol in order to calculate the frequency-dependent backscatter coefficient in BMF (Wear, 1999). Meanwhile, conventional ultrasound color Doppler images of the BMF solution pumped through a wall-less vessel embedded in a HIFU tissue mimicking phantom (King *et al.*, 2007) were evaluated using a 5.3 MHz linear array ultrasound imaging probe (VF7-3, Antares, Siemens Medical Solutions, Mountain View, CA).

4. Thermal properties

The dominant physical mechanism for HIFU therapy is the conversion of acoustical energy into thermal energy so as to ablate biological tissues through thermal necrosis (ter Haar and Coussios, 2007). The thermal properties of targeted tissues, e.g., conductivity and diffusivity, play an important role in the development of the *in situ* temperature distribution and corresponding therapeutic effects. Therefore, the thermal conductivity and diffusivity of the BMF were quantified with a thermal property analyzer (KD-2, Decagon Devices Inc, Pullman, WA) from 20 to 70 °C. The published values of the thermal conductivity (κ, W m^{-1} $°C^{-1}$) and diffusivity (D, mm^2 s^{-1}) for human blood are typically 0.53 and 0.12, respectively (Duck, 1990). Glycerol was chosen as a calibration reference in this experiment.

5. Dynamic viscosity

The most important rheological property of human blood, dynamic viscosity (mPa S or cP) is a function of both hematocrit and shear rate at moderate flow level (<45 s^{-1}) (Eckmann *et al.*, 2000). For larger shear rates (>45 s^{-1}), the blood viscosity reaches a plateau level around 4 cP at body temperature (37 °C). This flow-thinning, non-Newtonian effect is thought to occur in small vessels, but it is reasonable to consider that blood is Newtonian in large vessels (Eckmann *et al.*, 2000). More importantly, the viscosity of human blood is also strongly temperature dependent, which is critical for HIFU experiments and simulations. The non-Newtonian behavior of the BMF was first evaluated using a computer-controlled rotational rheometer (AR-G2, TA Instruments, New Castle, DE) across a wide range of flow shear rates. Then the temperature-dependent dynamic viscosity of the BMF was quantified with another rotational DV-E viscometer (Brookfield Inc., Middleboro, MA) from 20 to 70 °C.

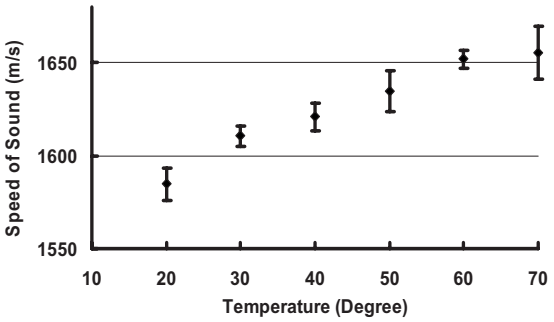


FIG. 2. Temperature-dependent speed of sound.

C. Statistical analysis

The means and standard deviations of the measured values for each BMF property were computed from at least three samples using Excel (Office XP, Microsoft Inc., Seattle, WA) and compared to the published values of human blood.

III. RESULTS

A. Speed of sound and ultrasound attenuation

Figure 2 shows the temperature-dependent sound speed in the BMF. Specifically, sound speed increased from 1585 m s^{-1} at 20 °C to 1621 m s^{-1} at 40 °C to 1655 m s^{-1} at 70 °C. This trend is comparable with that typically observed in biological tissues (Duck, 1990).

The frequency-dependent attenuation coefficient was characterized through power law ($\alpha = a \cdot f^b$) regression ($R > 0.95$) from 2 to 8 MHz based on experimental results to yield the specific coefficients a and b. Figure 3(a) is the ultrasound attenuation of the BMF at 20 °C. The corresponding power law regression was found to be $0.2f^{1.24}$ dB cm^{-1}, which is very close to the reported attenuation in human blood (Duck, 1990). Figure 3(b) presents the coefficient a as a function of temperature from 20 to 70 °C. An increase in a with temperature was found, the values at 20, 40, and 70 °C being 0.20, 0.23, and 0.27, respectively. On the other hand, the exponential coefficient b of the power law equation dropped from 1.21 to 1.03 as the sample temperature increased [Fig. 3(c)]. Combining the results for both coefficients a and b, it is interesting that the overall frequency-dependent ultrasound attenuation of the BMF increases by 20% at 2 MHz while it decreases by 7% at 8 MHz as the temperature increases from 20 to 70 °C.

B. Nonlinear parameter *B/A*

At room temperature (20 °C), the B/A of the BMF was measured to be 5.6 ± 0.6, which is close to that of human blood (Duck, 1990).

C. Backscatter coefficient and color Doppler image

Figure 4(a) depicts the backscatter coefficient as a function of frequency from 2 to 5 MHz at room temperature. Within this frequency range, the BMF has comparable backscatter coefficient levels with human blood (Nassiri and Hill, 1986; Yuan and Shung, 1988). For instance, the averaged

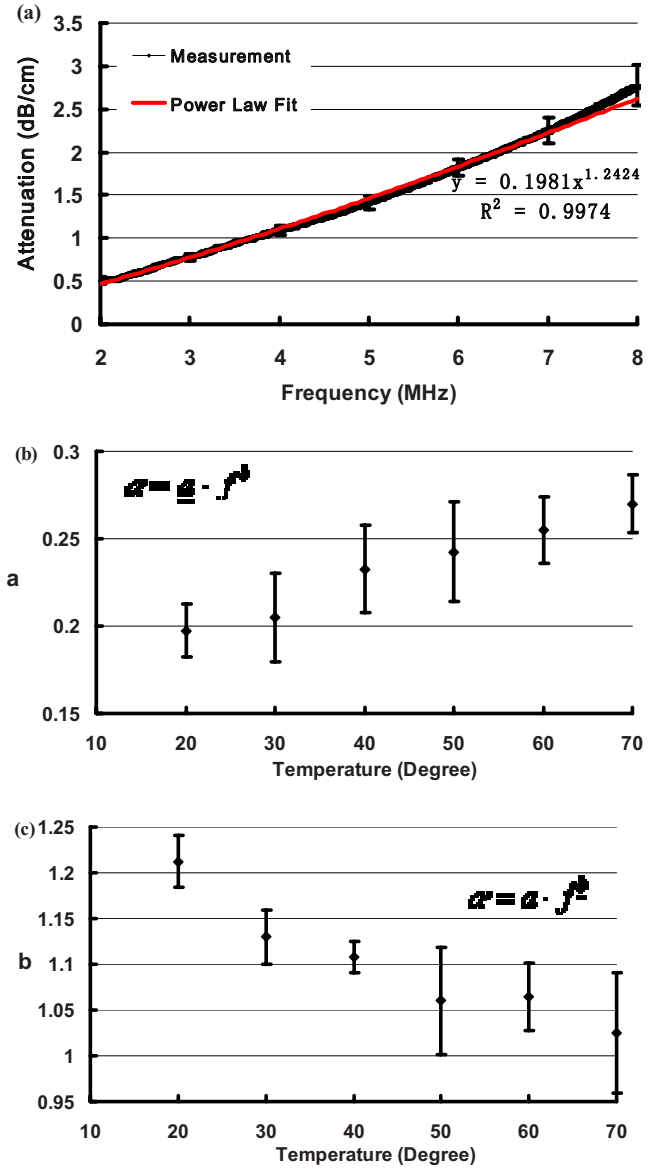


FIG. 3. (Color online) (a) Frequency-dependent ultrasound attenuation at 20 °C. Temperature-dependent (b) attenuation coefficient slope a and (c) attenuation coefficient exponential power b.

backscatter coefficient of the BMF is $5.3 \pm 2.6 \times 10^{-5}$ cm^{-1} Sr^{-1} at 3.5 MHz, while that of human blood is about $(2-5) \times 10^{-5}$ cm^{-1} Sr^{-1} at this frequency (Duck, 1990). Since the particle size is much less than the wavelength (430 μm) of the interrogating frequency, power law fitting reveals a quartic ($n=4.1$) frequency-dependent backscatter coefficient indicating the Rayleigh scattering nature of the BMF (Nassiri and Hill, 1986). A Doppler ultrasound image from the BMF in a HIFU flow phantom is presented in Fig. 4(b). Clear Doppler signature and backscattering contrast were detected in the ultrasound image without major clumps in the flow.

D. Thermal properties

Using glycerol as a standard reference, good agreement was obtained between the thermal property analyzer measurement ($\kappa = 0.28$ W m^{-1} °C^{-1}, $D = 0.10$ mm^2 s^{-1}) and pub-

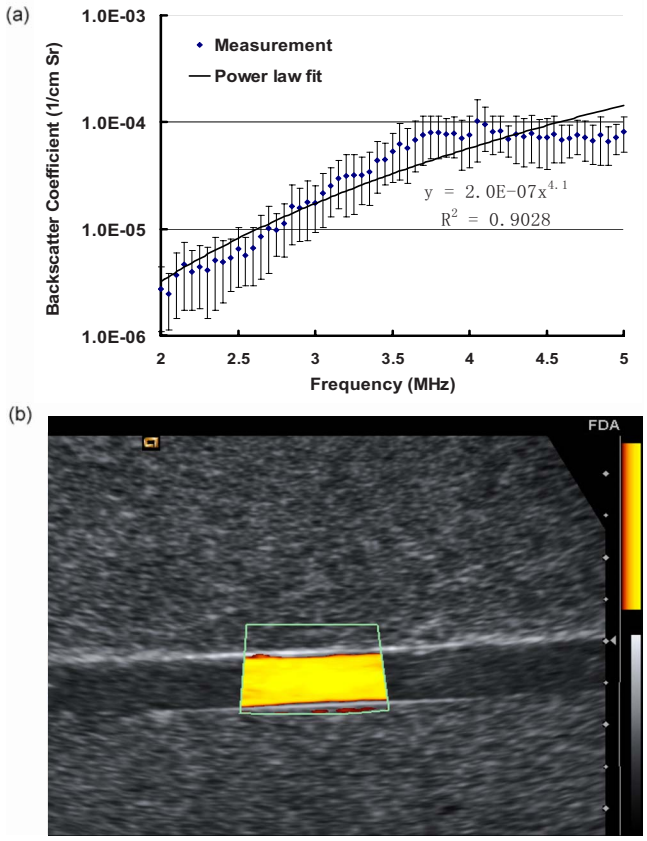


FIG. 4. (Color online) Measurement of (a) frequency-dependent backscatter coefficient and (b) Doppler ultrasound image.

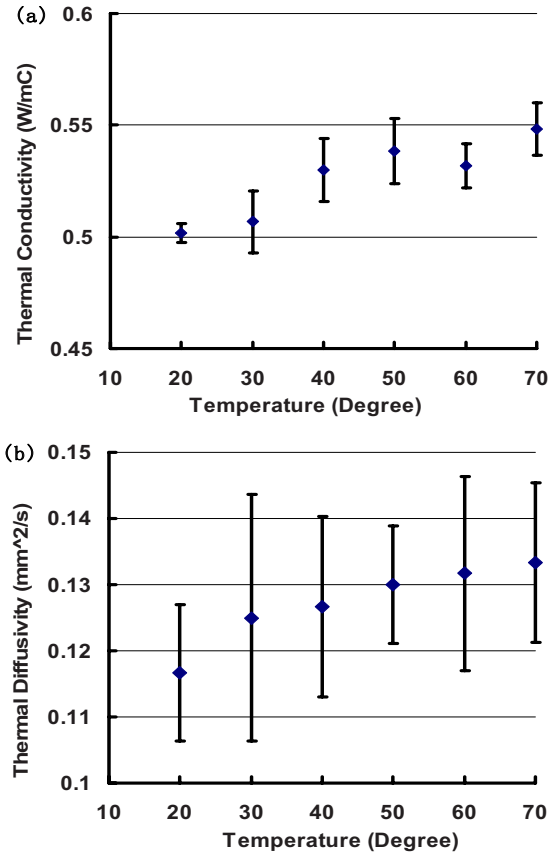


FIG. 5. (Color online) Temperature-dependent of (a) thermal conductivity and (b) thermal diffusivity.

lished values (Lide, 2004). The thermal conductivity (κ) and diffusivity (D) for the BMF are plotted as a function of temperature in Fig. 5. The measurement results support that the BMF (κ=0.5 W m^{-1} °C^{-1}, D=0.12 mm^2 s^{-1}) at room temperature (20 °C) has very close thermal properties to human blood (κ=0.5–0.53 W m^{-1} °C^{-1}, D=0.12 mm^2 s^{-1}) at body temperature (37 °C). Importantly, both of the thermal parameters augmented slightly (8% for thermal conductivity and 18% for thermal diffusivity) from 20 to 70 °C, which is also concordant with general observations in soft tissues (Duck, 1990). The thermal property measurements demonstrate the potential application of the BMF in HIFU therapy studies.

E. Dynamic viscosity

Figure 6(a) reveals an initial decreasing viscosity at extremely low shear rate ($<1/s$) and a plateau viscosity level (3.8 cP) over the entire shear rate range of 1–100/s. This result is different from the non-Newtonian behavior observed in the human blood (Eckmann *et al.*, 2000) and therefore the current BMF should be considered as a Newtonian fluid. However, human blood is also considered Newtonian for flow in large vessels (shear rate >45/s), which is more relevant for most HIFU studies. The viscosity was measured to be about 3.8 cP at room temperature (20 °C), which is close to the reported value for human blood at body temperature. As the temperature increases, the results demonstrate a steady decrease from 3.8 to 3 cP to 1.7 cP at 20, 40, and 70 °C, respectively [Fig. 6(b)].

F. Summary comparison with human blood

Table II summarizes all the measured BMF physical properties at 20 °C in comparison with that of human blood

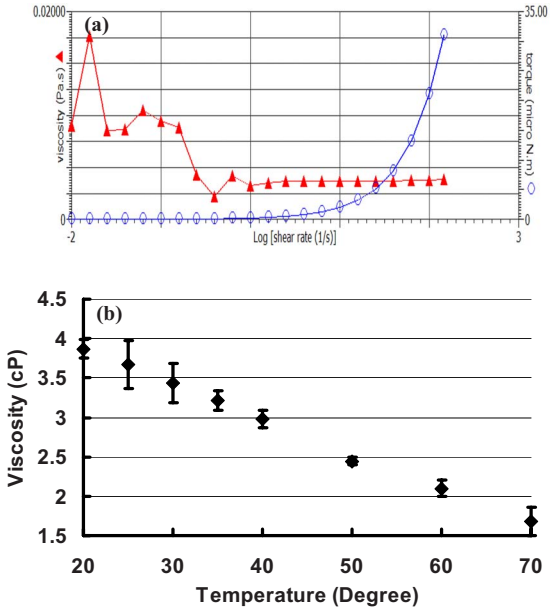


FIG. 6. (Color online) (a) Shear rate and (b) temperature dependent dynamic viscosity.

TABLE II. Comparison between BMF and Human Blood (20 °C).

Physical properties	HIFU BMF	Human Blood (Duck, 1990)
Viscosity, cP (mPa s)	3.8	3–4 at 37 °C
Thermal diffusivity, $mm^2\ s^{-1}$	0.11 ± 0.1	0.1–0.12
Thermal conductivity, $W\ m^{-1}\ °C^{-1}$	0.50	0.5–0.58
Acoustic attenuation, $dB \cdot cm^{-1}$	$0.24f^{1.24}$	$(0.16–0.24)f^{(1.19–1.23)}$
Speed of sound, $m\ s^{-1}$	1585 ± 17	1540–1590
Nonlinear parameter B/A	5.6 ± 0.6	6.3
Backscatter coefficient, $cm^{-1}\ Sr^{-1}$	$5.3 \pm 2.6 \times 10^{-5}$ at 3.5 MHz	$(2–5) \times 10^{-5}$ at 3.5 MHz

available in the literature (Duck, 1990; Samavat and Evans, 2006). Due to the biological heterogeneity, even the reported values of human blood vary within a certain range. However, it is noticeable that all the BMF properties are within the general levels of human blood. Most importantly, the BMF is the first reported phantom possessing attenuation and nonlinear parameter B/A properties similar to those of human blood, which may be beneficial for the HIFU application. The denaturation process in human blood during HIFU ablation was not intended to be mimicked considering the device calibration purpose of the BMF.

IV. DISCUSSION AND CONCLUSION

The aim of this study was to develop a BMF for safety and effectiveness evaluations of HIFU ablation devices. A HIFU BMF consisting of LDPE, nylon, gellan gum, and glycerol was devised and characterized. Compared to conventional ultrasound Doppler flow phantoms, a HIFU BMF must possess a range of appropriate acoustical and thermal properties (not just the backscatter coefficient) for valid measurements and accurate comparisons to theoretical modeling of HIFU exposures. Therefore, acoustical properties such as ultrasound attenuation, speed of sound, nonlinear parameter B/A, backscatter coefficient, viscosity, and also thermal properties were evaluated over a temperature range of 20–70 °C. Although the available data describing these blood properties are incomplete in the literature, especially at elevated temperatures where thermal coagulation occurs, the measured physical properties of the BMF were comparable to the values that have been reported for human blood (Duck, 1990; Eckmann *et al.*, 2000). It is notable that all the additives have high thermal stability (>95 °C), which is essential for a HIFU BMF considering the high temperature applications.

The ultrasound attenuation of the BMF was found to be largely dictated by the two primary additives, nylon and LDPE microspheres. Nylon particles were dispersed in the solution to mainly enhance the Rayleigh scattering due to its significant acoustic impedance mismatch with the surrounding fluid. On the other hand, LDPE has close acoustic impedance to soft tissue and a linear frequency specific ultrasound absorption. Therefore, LDPE was found to be an excellent additive for enhancing the ultrasound absorption without altering the scattering property. When these two components were dispersed in the fluid simultaneously, the overall attenuation was approximately linear with frequency and the attenuation slope was controllable around 0.2 $dB\ cm^{-1}\ MHz^{-1}$ at room temperature. The backscatter coefficient was adjusted to about $5.3 \times 10^{-5}\ cm^{-1}\ Sr^{-1}$ at 3.5 MHz. Both attenuation and backscatter are similar to those of human blood at room temperature. Higher or lower attenuation and backscatter coefficients are attainable through optimizing the LDPE and nylon content ratio in the BMF. It also was noted that the density, nonlinear parameter B/A, thermal properties, and viscosity are mainly dependent on the content ratio of water, glycerol, and gellan gum powder. The primary advantage of such a synergistic combination of multiple controllable additives instead of any single one is that it provides an effective and flexible method for engineering BMFs with various desired physical properties for any specific application.

It is worthwhile to mention that the BMF density is mainly controlled by the glycerol content, which was fine-tuned to keep the nylon and LDPE particles largely suspended (although in actual phantom use, the stirring action of the fluid flow will accomplish this suspension). Special attention was paid to keep the particles suspended by rotating/shaking the phantom holder during all the measurements. The variation of sound speed, attenuation, and B/A observed in the measurements may be partially due to the nonuniform membrane surface in the sample holder and the accuracy of sample thickness measurements. In further studies, it will be beneficial to understand the underlying physical mechanism of the LDPE and gellan gum particle dispersions in determining the acoustical and thermal properties in the BMF.

The use of conventional B-mode imaging to monitor HIFU therapy typically depends on the appearance of hyperechoic regions in the HIFU focus due to the bubble formation (Coussios *et al.*, 2007). The backscatter coefficient and Doppler imaging measurements demonstrated the potential application of the BMF for evaluating hemostasis therapy devices. It should be noted that the backscatter coefficient of the HIFU tissue mimicking material used for Doppler imaging [Fig. 4(b)] is much lower (>20 dB) than that of biological tissues. Therefore, the contrast between the BMF and the background material in Fig. 4(b) is not as high as real blood in soft tissues (~40 dB).

Having similar acoustical, thermal, and mechanical properties to real blood, the BMF opens up the possibility of developing a well-controlled ultrasound flow phantom for HIFU thermal ablation research. Multiple interconnected

wall-less flow channels mimicking different types of blood vessels (artery, vein, capillary, etc.) and bleeding can be tunneled through a tissue mimicking phantom with embedded thin-wire thermocouples surrounding the potential HIFU focus. With the BMF circulating in this phantom system, the effect of blood flow, including vessel diameter, flow rate, thermal convection, nonlinear waveform distortion, and cavitation-enhanced absorption and shielding, can be investigated systemically. Furthermore, the temperature-dependent properties reported in this work will enable more accurate and comprehensive numerical simulations, which can be subsequentially compared and correlated with the focal temperature distribution measured from the HIFU flow phantom system.

In summary, this BMF devised for HIFU applications is a degassed water solution dispersed with LDPE, nylon, gellan gum, and glycerol. The temperature dependence of the acoustical, thermal, and mechanical properties were studied from 20 to 70 °C based on established methods. With similar physical properties to human blood, this BMF provides a potential tool for developing standardized exposimetry techniques, validating numerical models, and determining the safety and efficacy of HIFU ablation devices.

ACKNOWLEDGMENTS

This research was supported by Defense Advanced Research Projects Agency (DARPA) through IAG No. 224-05-6016. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the U.S. Department of Health and Human Services.

APPENDIX: BLOOD MIMICKING FLUID PROTOCOL

The steps for BMF are as follows.

(1) Add 100 ml of degassed and deonized water to an empty glass beaker.
(2) Start stirring the water with a magnetic stirring hotplate (11-500-49SH, Fisher Scientific, Hampton, NH) and dissolve 0.1 g gellan gum powder in the cold water completely and uniformly.
(3) Start heating the above gellan gum solution and keep stirring gently to avoid bubble agitation.
(4) Boil the solution for 2–3 min until it turns clear and cool at room temperature.
(5) Once at room temperature, add 10 ml glycerol and 1 ml surfactant into the solution.
(6) Add 0.5 g nylon particles and 6 g LDPE powder into the solution.
(7) Keep stirring the solution gently until all additives disperse uniformly.
(8) After the mixing process, pour the fluid slowly into holders for measurements.

Bailey, M. R., Khokhlova, V. A., Sapozhnikov, O. A., Kargl, S. G., and Crum, L. A. (**2003**). "Physical mechanisms of the therapeutic effect of ultrasound," Acoust. Phys. **49**, 437–464.

Boote, E. J., and Zagzebski, J. A. (**1988**). "Performance tests of Doppler ultrasound equipment with a tissue and blood mimicking phantom," J. Ultrasound Med. **7**, 137–47.

Chapelon, J. Y., Ribault, M., Vernier, F., Souchon, R., and Gelet, A. (**1999**). "Treatment of localized prostate cancer with transrectal high intensity focused ultrasound," Eur. J. Ultrasound **9**, 31–38.

Chaussy, C., and Thuroff, S. (**2003**). "The status of high-intensity focused ultrasound in the treatment of localized prostate cancer and the impact of a combined resection," Curr. Urol. Rep. **4**, 248–252.

Chen, L., ter Haar, G., Hill, C. R., Dworkin, M., Carnochan, P., Young, H., and Bensted, J. P. (**1993**). "Effect of blood perfusion on the ablation of liver parenchyma with high-intensity focused ultrasound," Phys. Med. Biol. **38**, 1661–1673.

Coussios, C. C., Farny, C. H., ter Haar, G. R., and Roy, R. A. (**2007**). "Role of acoustic cavitation in the delivery and monitoring of cancer treatment by high-intensity focused ultrasound," Int. J. Hyperthermia **23**, 105–120.

Dong, F., Madsen, E. L., MacDonald, M. C., and Zagzebski, J. A. (**1999**). "Nonlinearity parameter for tissue-mimicking materials," Ultrasound Med. Biol. **25**, 831–838.

Duck, F. A. (**1990**). *Physical Properties of Tissue: A Complete Reference Book* (Academic, London).

Eckmann, D. M., Bowers, S., Stecker, M., and Cheung, A. (**2000**). "Hematocrit, volume expander, temperature, and shear rate effects on blood viscosity," Anesth. Analg. (Baltimore) **91**, 539–545.

Filonenko, E. A. and V. A. Khokhlova (**2001**). "Effect of acoustic nonlinearity on heating of biological tissue by high-intensity focused ultrasound," Acoust. Phys. **47**, 468–475.

Gammell, P. M., Maruvada, S., and Harris, G. R. (**2007**). "An ultrasonic time delay spectrometry (TDS) system employing digital processing," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1036–1044.

Greaby, R., Zderic, V., and Vaezy, S. (**2007**). "Pulsatile flow phantom for ultrasound image-guided HIFU treatment of vascular injuries," Ultrasound Med. Biol. **33**, 1269–1276.

Hariharan, P., Myers, M. R., and Banerjee, R. K. (**2007**). "HIFU procedures at moderate intensities—effect of large blood vessels," Phys. Med. Biol. **52**, 3493–34513.

Harris, G. R., Liu, Y., Maruvada, S., and Gammell, P. M. (**2007**). "Finite amplitude method for measurement of nonlinearity parameter B/A using plane-wave tone bursts," Proc.-IEEE Ultrason. Symp..

Huang, J., Holt, R. G., Cleveland, R. O., and Roy, R. A. (**2004**). "Experimental validation of a tractable numerical model for focused ultrasound heating in flow-through tissue phantoms," J. Acoust. Soc. Am. **116**, 2451–2458.

Hynynen, K., Pomeroy, P., Smith, D. N., Huber, P. E., McDannold, N. J., Kettenbach, J., Baum, J., Singer, S., and Jolesz, F. (**2001**). "MR imaging-guided focused ultrasound surgery of fibroadenomas in the breast: a feasibility study," Radiology **219**, 176–185.

Ichihara, M., Sasaki, K., Umemura, S., Kushima, M., and Okai, T. (**2007**). "Blood flow occlusion via ultrasound image-guided high-intensity focused ultrasound and its effect on tissue perfusion," Ultrasound Med. Biol. **33**, 452–459.

Jolesz, F. A., K. Hynynen, N.McDannold, D. Freundlich, and D. Kopelman (**2004**). "Noninvasive thermal ablation of hepatocellular carcinoma by using magnetic resonance imaging-guided focused ultrasound," Gastroenterology **127**, S242–S247.

Kennedy, J. E., ter Haar, G. R., Wu, F., Gleeson, F. V., Roberts, I. S., Middleton, M. R., and Cranston, D. (**2004**). "Contrast-enhanced ultrasound assessment of tissue response to high-intensity focused ultrasound," Ultrasound Med. Biol. **30**, 851–854.

King, R. L., Herman, B. A., Maruvada, S., Wear, K. A., and Harris, G. R. (**2007**). "Development of a HIFU phantom," *Proceedings: Sixth International Symposium on Therapeutic Ultrasound*, edited by C. C. Coussios and G. ter Haar (American Institute of Physics, New York), pp. 351–356.

Law, W. K., Dunn, F., and Frizzell, L. A. (**1984**). "The nonlinearity parameter B/A of biological media," Proceedings of the Tenth Symposium on Nonlinear Acoustics (unpublished), pp. 222–224.

Lide, D. R. (**2004**). *CRC Handbook Chemistry and Physics* (CRC, Boca Raton, FL).

Madsen, E. L., Dong, F., Frank, G. R., Garra, B. S., Wear, K. A., Wilson, T., Zagzebski, J. A., Miller, H. L., Shung, K. K., Wang, S. H., Feleppa, E. J., Liu, T., O'Brien, W. D., Topp, K. A., Sanghvi, N. T., Zaitsev, A. V., Hall, T. J., Fowlkes, J. B., Kripfgans, O. D., and Miller, J. G. (**1999**). "Interlaboratory comparison of ultrasonic backscatter, attenuation, and speed measurements," J. Ultrasound Med. **18**, 615–631.

Marczak, W. (**1997**). "Water as a standard in the measurements of speed of sound in liquids," J. Acoust. Soc. Am. **102**, 2776–2779.

Martin, R. W., Vaezy, S., Kaczkowski, P. J., Keilman, G., Carter, S., Caps, M., Beach, K., Plett, M., and Crum, L. A. (**1999**). "Hemostasis of punctured vessels using Doppler-guided high-intensity ultrasound," Ultrasound Med. Biol. **25**, 985–990.

McDannold, N., Tempany, M., Fennessy, M., So, J., Rybicki, J., Stewart, A., Jolesz, A., and Hynynen, K. (**2006**). "Uterine leiomyomas: MR imaging-based thermometry and thermal dosimetry during focused ultrasound thermal ablation," Radiology **240**, 263–272.

Nassiri, D. K., and Hill, C. R. (**1986**). "The differential and total bulk acoustic scattering cross sections of some human and animal tissues," J. Acoust. Soc. Am. **79**, 2034–2047.

Oates, C. P. (**1991**). "Towards an ideal blood analogue for Doppler ultrasound phantoms," Phys. Med. Biol. **36**, 1433–1442.

O'Neill, T. P., Winkler, A. J., and Wu, J. (**1994**). "Ultrasound heating in a tissue-bone phantom," Ultrasound Med. Biol. **20**, 579–588.

Preston, R. C., Shaw, A., and Zeqiri, B. (**1991**). "Prediction of in situ exposure to ultrasound: An acoustical attenuation method," Ultrasound Med. Biol. **17**, 317–332.

Ramnarine, K. V., Nassiri, D. K., Hoskins, P. R., and Lubbers, J. (**1998**). "Validation of a new blood mimicking fluid for use in Doppler flow test objects," Ultrasound Med. Biol. **24**, 451–459.

Samavat, H., and Evans, J. A. (**2006**). "An ideal blood mimicking fluid for Doppler ultrasound phantoms," J. Math. Phys. **31**, 275–278.

ter Haar, G. R., and Coussios, C. C. (**2007**). "High intensity focused ultrasound: Physical principles and devices," Int. J. Hyperthermia **23**, 89–104.

Vaezy, S., and Zderic, V. (**2007**). "Hemorrhage control using high intensity focused ultrasound," Int. J. Hyperthermia **23**, 203–211.

Wear, K. A. (**1999**). "Frequency dependence of ultrasonic backscatter from human trabecular bone: theory and experiment," J. Acoust. Soc. Am. **106**, 3659–3664.

Wiart, M., Curiel, L., Gelet, A., Lyonnet, D., Chapelon, J. Y., and Rouvière, O. (**2007**). "Influence of perfusion on high-intensity focused ultrasound prostate ablation: A first-pass MRI study," Magn. Reson. Med. **58**, 119–127.

Yuan, Y., and Shung, K. (**1988**). "Ultrasonic backscatter from flowing whole blood II: dependence on frequency and fibrinogen concentration," J. Acoust. Soc. Am. **84**, 1195–1200.

Ultrasonic velocity dispersion in bovine cortical bone: An experimental study

Guillaume Haïat[a)]
CNRS, Laboratoire de Recherches Orthopédiques, UMR CNRS 7052 B2OA, 10 avenue de Verdun, 75010 Paris, France

Magali Sasso and Salah Naili
Laboratoire de Mécanique Physique, Université Paris-Est, UMR CNRS 7052 B2OA, 61, avenue du Général de Gaulle, F-94010 Créteil, France

Mami Matsukawa
Laboratory of Ultrasonic Electronics, Faculty of Engineering, Doshisha University, Kyotanabe, Kyoto-fu 610-0321, Japan

(Received 4 December 2007; revised 26 May 2008; accepted 31 May 2008)

Cortical bone quality is determinant in bone fragility and its ultrasonic evaluation has become possible in clinical practice. However, the interaction between a broadband ultrasonic pulse and this complex multiscale medium remains poorly understood. The frequency dependence of phase velocity, which may impact clinical measurements, has been sparsely investigated. Our objective is to evaluate the determinants of the frequency dependence of phase velocity in bovine femoral cortical bone samples using an *in vitro* ultrasonic transmission device. The apparent phase velocity varies quasilinearly on a 1 MHz restricted bandwidth around 4 MHz. After compensating for diffraction effects, significant differences in velocity dispersion are obtained according to the anatomical location. The microstructure of each sample is determined using an optical microscope, which allows assessing the dependence of dispersion on the type of bone microstructure. Mostly positive but also negative values of dispersion are measured. Negative dispersion is mainly obtained in samples constituted of mixed microstructure, which may be explained by phase cancellation effects due to the presence of different microstructures within the same sample. Dispersion is shown to be related to broadband ultrasonic attenuation values, especially in the radial direction. Results are compared with the local Kramers–Kronig relationships. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2950091]

PACS number(s): 43.80.Ev, 43.80.Vj, 43.35.Cg, 43.20.Jr [FD] Pages: 1811–1821

I. INTRODUCTION

Bone quantitative ultrasound (QUS) techniques are now widely used for the assessment of osteoporosis and fracture risk prediction. The gold standard for the diagnosis of osteoporosis (Cullum *et al.*, 1989) is dual-energy x-ray absorptiometry (DXA) techniques, which give access to information related to bone mass: bone mineral density (BMD). BMD is known to be related to fracture risk (Kanis, 2002). However, other bone characteristics such as microarchitecture or material properties may also play a role in bone strength and are difficult to obtain with DXA. As an elastic wave, ultrasound has the ability not only to assess bone mass (Haïat *et al.*, 2007b) but also to go beyond this estimation and provide additional information on bone mechanical quality (Haiat *et al.*, 2008a).

Currently, the development of QUS techniques is still limited (Cadossi *et al.*, 2000) since the information potentially available in the transmitted ultrasonic wave is not fully analyzed and parameters such as bone material properties or microstructural parameters may still be difficult to recover. The interaction between ultrasound and bone remains poorly understood from a physical point of view due to the complex nature of bone which has a viscoelastic porous microstructure spanning many length scales (Cowin, 2001). Cortical bone is a dispersive medium in which relatively important values of attenuation coefficients have been measured [e.g., around 40 dB cm^{-1} at 4 MHz (Sasso *et al.*, 2007)].

Initially, most applications of QUS techniques to bone have been confined to cancellous bone characterization (Langton *et al.*, 1984; Laugier *et al.*, 1994; Nicholson and Njeh, 1999). However, interest should also be placed in cortical bone exploration (Rico, 1997) because (i) it supports most of the load of the body and (ii) it is involved in many osteoporotic fractures. Furthermore, cortical bone is affected by age-related bone resorption and osteoporosis (Cowin, 2001).

The development of new QUS devices such as the axial transmission technique currently enables the *in vivo* evaluation of cortical bone. This technique allows the assessment of the cortical layer of different anatomical sites (Foldes *et al.*, 1995; Barkmann *et al.*, 2000; Bossy *et al.*, 2004a). QUS techniques for cortical bone evaluation are mainly confined to speed of sound (SOS) analysis and numerous *in vitro* studies (see, for example, Ashman *et al.*, 1984; Lee *et al.*, 1997;

a)Author to whom correspondence should be addressed. Tel.: (+33) 1 44 89 77 64. FAX: (+33) 1 44 89 78 22.

Bensamoun *et al.*, 2004a; Bensamoun *et al.*, 2004b) focused on ultrasonic wave propagation in cortical bone and on its relation to bone physical parameters through wave velocity analysis. In their pioneering work, Lees *et al.* (Lees *et al.*, 1979; Lees *et al.*, 1983) showed that the sonic plesiovelocity in cortical bone depends significantly on the direction of propagation as well as on bone properties. However, in comparison with trabecular bone studies, few investigations (Yoon and Katz, 1976; Lakes *et al.*, 1986; Lees and Klopholz, 1992) have reported measurements of the frequency dependence of ultrasonic phase velocity in cortical bone. In addition, no systematic study has been found in literature about the frequency dependence of phase velocity in cortical bone.

Velocity dispersion is an important parameter since it affects SOS measurements. There are many different ways of measuring SOS in the laboratory and in the clinics. First, using heel transverse transmission devices, the ultrasonic velocity can be measured as the phase or the group velocity; the latter quantity being a function of dispersion, frequency, and phase velocity (Morse and Ingard, 1986; Duck, 1990). Moreover, SOS measurements deduced from "early" time-of-flight measurements of an ultrasonic pulse transmitted through bone have been shown to be the optimal method for the assessment of BMD estimated with DXA techniques in intact human proximal femur (Haiat *et al.*, 2005). Wear (Wear, 2000b; Wear, 2001a) reported on differences in methods for measuring ultrasonic velocity in bone, including phase and group velocities as well as transit-time-based SOS estimates. Time-of-flight measurements are subject to bias due to the modification of the pulse shape during propagation through bone by frequency-dependent attenuation (Wear, 2000b; Wear, 2001a) and dispersion (Haïat *et al.*, 2006; Wear, 2007b) producing artifacts in SOS measurements. Second, using axial transmission devices, the ultrasonic velocity may also be measured using different signal processing techniques such as time markers (first arriving signal) (Bossy *et al.*, 2004b), two dimensional fast Fourier transform analysis (phase velocity) (Lefebvre *et al.*, 2002; Moilanen *et al.*, 2006), time frequency analysis (group velocity) (Protopappas *et al.*, 2006; Ta *et al.*, 2006), or wave extraction techniques (Sasso *et al.*, 2007). In axial transmission, the velocity dispersion measured is influenced by the geometry (e.g., guided wave effects) as well as by phenomena occurring in the bulk of the material. Note that these last phenomena also occur in trabecular bone in the context of transverse transmission devices. For all these reasons, dispersion is an important acoustical property, influencing significantly any velocity measurements. A better understanding of its physical determinant may have important impacts on the clinical relevance of SOS measurements.

The present study focuses on the evaluation of the frequency-dependent phase velocity in bovine cortical bone. The first objective of this study is to evaluate the feasibility of dispersion measurements around 4 MHz. This range of frequency yields a wavelength in cortical bone around 1 mm, which is higher but comparable with the typical size of the osteonal structures (100 μm) (Currey, 1984; Locke, 2004), thus allowing the ultrasound wave to be sensitive to bone structural properties (Sasso *et al.*, 2008). The second objective is to investigate the dependence of velocity dispersion on the anatomical position and on the direction of propagation relative to the bone axis. The third objective is to investigate the relationship between dispersion, bone microstructural properties, and bone mass density.

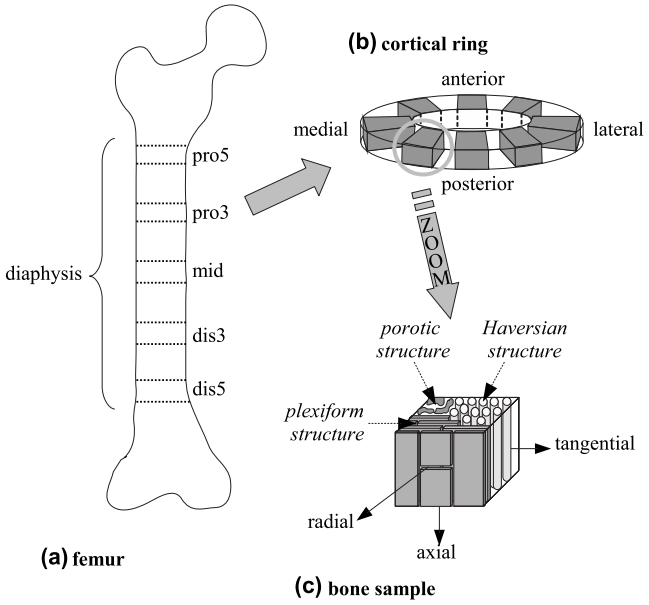


FIG. 1. Schematic representation of the spatial distribution of the bone samples. (a) Locations where the intact femur was cut to obtain the five cortical rings. (b) Quadrant positions of the eight parallelepipedic samples around each cortical ring. (c) Illustration of the orientation of the three directions and of the three types of microstructures.

In this paper, parallelepipedic cortical bone samples were analyzed using a multimodal approach. The microstructure of all the samples was assessed by optical microscopy and their apparent volumetric bone mineral density (vBMD) was measured using a DXA technique. The samples were then analyzed using an ultrasonic transmission device in the axial, radial, and tangential directions relative to the femur orientation.

II. MATERIALS AND METHODS

This section briefly summarizes the experiments described in greater detail in several papers (Yamato *et al.*, 2006; Sasso *et al.*, 2007; Sasso *et al.*, 2008; Yamato *et al.*, 2008).

A. Sample preparation

Three intact left femurs from 36-month old bovines were obtained at a public abattoir. Figure 1 schematically shows how the bovine femurs were cut. From each femur, five ring shaped cortical bones were cut (i) in the distal part at 50 and 30 mm from the center of the bone (termed, respectively, dis5 and dis3), (ii) at the center of the bone (termed mid), and (iii) in the proximal part at 30 and 50 mm from the center of the bone (termed, respectively, pro3 and pro5). Eight parallelepipedic samples were cut from each ring in the anterior, posterior, medial, lateral, and in the four oblique parts. The samples were polished into rectangular

parallelepipeds with sides comprised between 4 and 11 mm. They were stored at a temperature of −20 °C and unfrozen prior to the different measurements.

B. Mass density and apparent volumetric bone mineral density measurements

For each sample, mass density was determined using Archimedes' principle. In addition, apparent vBMD was determined with a clinical DXA device (QRD-1000, Hologic, Bedford, MA). BMD data were converted into vBMD by taking into account the size of each sample in the axial direction.

C. Analysis of the microstructure

Bone sections were cut from the top surface (in the axial direction) of each sample and analyzed using an optical microscope. Following a previous study (Sasso *et al.*, 2008), the microstructure of each bovine cortical bone sample was manually classified into four distinct structures termed plexiform (Pl), Haversian (H), porotic (Po), and mixed (M). The last microstructure corresponds to samples composed of different microstructures. The porotic microstructure differs from the other ones by a larger pore size (50–300 μm) compared with pores sizes in Haversian (20–50 μm) and plexiform structures (8–12 m). 47% of the samples had a plexiform microstructure, 19% a Haversian microstructure, 8% a porotic microstructure, and 26% a mixed microstructure.

D. Ultrasonic measurements

The ultrasonic transmission setup has been described in detail in previous publications (Sasso *et al.*, 2007; Sasso *et al.*, 2008) and is briefly recalled in what follows. Ultrasonic measurements were performed using a classical substitution method. A pair of self-made broadband polyvinylidene fluoride (PVDF) planar transducers with an 8 mm diameter was used. The PVDF transducers have been described in Nakamura and Otani, 1993 and have the advantage of generating short pulses, as they are effective on a wide bandwidth. Their resolution in the time domain allows a better understanding of wave propagation phenomena. The emitter and receiver were coaxially aligned and operated in transmission. The transducer interdistance was equal to 21.2 mm. The emitter was driven by a function generator 33250A manufactured by Agilent Technologies (Palo Alto, CA). Received signals were amplified using a wideband amplifier BX-31A from NF Electronic Instruments (Yokohama, Japan) and digitized by a digital oscilloscope. Each received signal was then transferred to a personal computer for off-line analysis. A broadband ultrasonic pulse (−10 dB bandwidth: 2.4–10.9 MHz) (Sasso *et al.*, 2007) was received first without and then with the sample positioned between the transducers. Samples were immersed in a normal saline solution (NSS) whose temperature is controlled [(25.0 ± 0.1) °C] and were degassed carefully during more than 1 h prior to ultrasonic measurements. High frequency noise was filtered using a low pass filter with a 20 MHz cutoff frequency. In order to recover the signal corresponding to the wave propagating within the bone sample only and to suppress possible effects of a wave propagating in water (along the sample), the received signals were windowed around the first received contribution using a time window whose length is much larger than the signal duration, so that dispersion measurements are not affected by the windowing procedure. Each sample was tested in the axial, radial, and tangential directions, as shown in Fig. 1.

The apparent phase velocity in bone $V_b(f)$ was estimated for each sample and each direction from the difference $\varphi(f)$ between the phase of the reference signal transmitted in NSS and the phase of the signal transmitted through the bone sample. The phase of each signal was evaluated as the argument of the spectrum obtained using a fast Fourier transform. This phase difference was unwrapped as described in Verhoef *et al.*, 1985 and Droin *et al.*, 1998. Phase unwrapping and phase sheet offset are critical issues for calculating phase velocity. The apparent phase velocity is then given by

$$V_b(f) = \frac{1}{\frac{1}{V_r} - \frac{\varphi(f)}{2\pi f L}}, \tag{1}$$

where L is the sample size in the direction of propagation, f is the frequency, and V_r is the wave velocity in NSS, which was assumed to be independent of the frequency. The velocity V_r was derived from the time-of-flight difference Δt between the ultrasonic pulse received in distilled water and in NSS using

$$V_r = \frac{V_w H}{H - \Delta t V_w}, \tag{2}$$

where H is the intersensor distance (21.2 mm) and V_w is the temperature-dependent wave velocity in distilled water given by (Kaye and Laby, 1973)

$$V_w = 1402.9 + 4.835T - 0.047\,016T^2 - 0.000\,127\,25T^3. \tag{3}$$

In Eq. (3), T is the temperature in °C. At 25 °C, $V_w = 1496.4$ m s^{-1} and $V_r = 1508$ m s^{-1}.

The computation of the apparent phase velocity may lead to biased estimate of the intrinsic phase velocity due to diffraction effects occurring between the two identical transducers (Kino, 1987). A set of approximate corrections can be used in insertion techniques to relate observed experimental signals to the phase velocity (Xu and Kaufman, 1993; Kaufman *et al.*, 1995). This method has been applied by Droin *et al.* (1998), so only the basic relations are noted here. The analysis is based on the excitation of longitudinal waves in a liquid medium by a finite circular piston source in an infinite rigid baffle radiating into a semi-infinite medium. In this method, the corrected phase difference $\varphi^c(f)$ used to compute the corrected phase velocity in bone $V_b^c(f)$ is obtained from the uncorrected phase difference $\varphi(f)$ through (Xu and Kaufman, 1993; Kaufman *et al.*, 1995; Droin *et al.*, 1998)

$$\varphi^c(f) = \varphi(f) - \arg\left(\frac{\int_0^\infty \frac{J_1^2(Y)}{Y} e^{jY^2 \frac{S}{4\pi}} dY}{\int_0^\infty \frac{J_1^2(Y)}{Y} e^{jY^2 \frac{S_w}{4\pi}} dY} \right), \quad (4)$$

where J_1 is the first order Bessel function of order zero. The quantities $S_w = HV_w/a^2 f$ and

$$S = \frac{[(H-L)V_w + LV_b(4\ \text{MHz})]}{a^2 f}$$

are, respectively, the Fresnel parameter for the water and water-sample-water paths, where a is the transducer radius (4 mm). The corrected phase velocity $V_b^c(f)$ was obtained using Eq. (1), with $\varphi^c(f)$ instead of $\varphi(f)$. Eventually, the velocity dispersion was obtained from the difference between the values of the corrected phase velocity at 4.5 and 3.5 MHz $[V_b^c(4.5\ \text{MHz}) - V_b^c(3.5\ \text{MHz})]$.

E. Short-term reproducibility

In order to evaluate the reproducibility of velocity dispersion measurements, four measurements were performed with repositioning for each sample and in each direction. The short-term reproducibility was evaluated on dispersion measurements as described hereafter.

The individual short-term precision SD_n^d was evaluated as the standard deviation (SD) of the four measurements for the sample *#n* in the direction *d*. Then, the precision error of the technique for a given direction was determined by the root-mean-square average of the individual precision errors using the procedure described by (Glüer *et al.*, 1995; Njeh *et al.*, 1999)

$$P^d = \sqrt{\sum_{n=1}^{N} \frac{\text{SD}_n^{d^2}}{N}}, \quad (5)$$

where N is the total number of samples.

F. Statistical analysis

Analysis of variance (ANOVA) and Tuckey–Kramer multiple comparison tests were performed using MATLAB software (The MathWorks Inc., Natick, MA) to test the significance of dispersion variations as a function of the measurement direction, anatomical location, and microstructure.

III. RESULTS

The apparent phase velocity $V_b(4\ \text{MHz})$ in bone was equal to $4116 \pm 176\ \text{m s}^{-1}$ in the axial direction, $3305 \pm 199\ \text{m s}^{-1}$ in the radial direction, and $3548 \pm 172\ \text{m s}^{-1}$ in the tangential direction. The magnitude of the correction from diffraction effects depends on the size of the sample considered as well as on the value of the apparent phase velocity at 4 MHz. For instance, when the apparent phase velocity at 4 MHz is equal to $3500\ \text{m s}^{-1}$, the correction of diffraction effects induces an increase in dispersion which is comprised between 1 and $4\ \text{m s}^{-1}$.

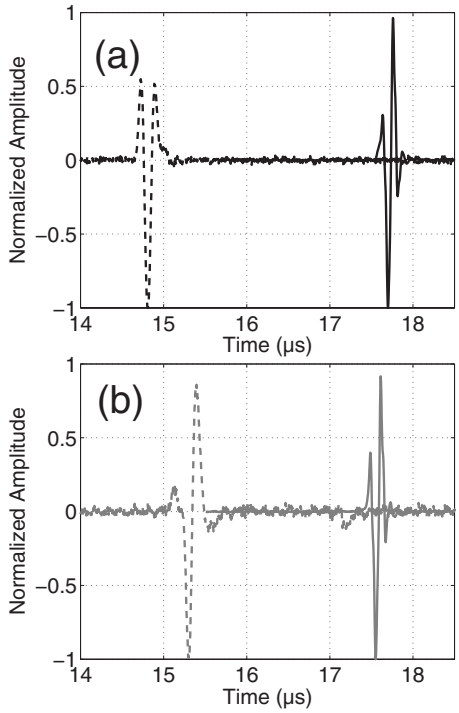


FIG. 2. Example of normalized rf signals. (a) The black dashed line represents a typical wave form obtained through a bone sample associated with a positive value of the dispersion. The ratio of the maximum amplitude of the signal transmitted through bone and through water is of 0.09. (b) The gray dashed line represents a typical wave form obtained through a bone sample associated with a negative value of the dispersion. The ratio of the maximum amplitude of the signal transmitted through bone and through water is of 0.04. The negative dispersion sample was taken from the anterodistal part at 5 cm from the center on the diaphysis; the direction of propagation was tangential. The positive dispersion sample was taken from the posteroproximal part at 3 cm from the center on the diaphysis; the direction of propagation was axial. In both figures, the reference wave form is obtained through NSS (solid lines).

A typical wave form obtained through a bone sample with a positive corrected dispersion value of $20\ \text{m s}^{-1}$ (negative corrected dispersion value of $-10.3\ \text{m s}^{-1}$, respectively) is plotted in Fig. 2(a) [Fig. 2(b), respectively] with a black (gray, respectively) dashed line. In Fig. 2, the solid lines correspond to the reference wave form obtained through NSS. The frequency dependence of phase velocity corresponding to both cases shown in Fig. 2 is plotted in Fig. 3 between 3.5 and 4.5 MHz.

In spite of a nonlinear variation of the apparent phase velocity over the whole frequency range (data not shown), the frequency dependence of the apparent phase velocity is quasilinear over the 1 MHz restricted frequency range around 4 MHz, for all samples in the three directions [determination coefficients (R^2) ranging from 0.96 to 1].

Precision of dispersion measurements in the axial (radial and tangential, respectively) direction is equal to $1.3\ \text{m s}^{-1}$ (respectively, 0.8 and $1.3\ \text{m s}^{-1}$). The standard deviation of the distribution of the individual short-term precision SD_n^d is equal to $1.1\ \text{m s}^{-1}$ in the axial direction (0.8 and $1.4\ \text{m s}^{-1}$ in the radial and tangential directions, respectively).

Average, standard deviation, and minimal and maximal values of dispersion obtained for the three directions of propagation are summarized in Table I. In average, disper-

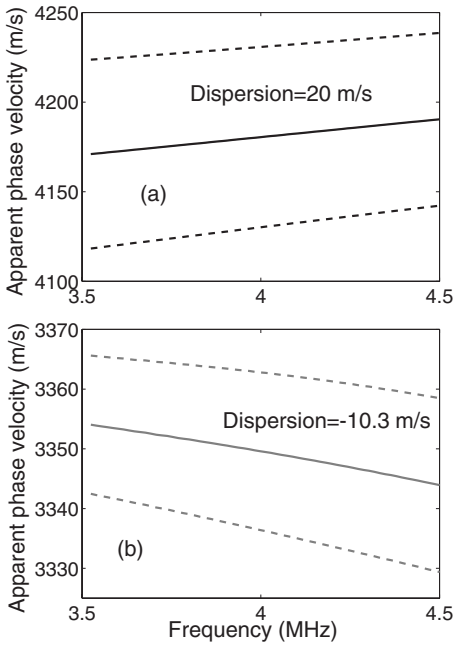


FIG. 3. Frequency dependence of the apparent phase velocity. (a) The black solid line shows the phase velocity obtained for the sample corresponding to the rf signal showed in Fig. 2(a) (positive dispersion). (b) The gray solid line shows the phase velocity obtained for the sample corresponding to the rf signal showed in Fig. 2(b) (negative dispersion). The dashed lines indicate the sum (respectively, the difference) of the apparent phase velocity and of the standard deviation obtained for the four measurements performed for the corresponding sample.

sion values obtained in the axial direction are higher than the ones obtained in the radial and tangential directions. However, the ANOVA test does not reveal a significant directional effect ($p=0.2$, $F=1.6$).

To assess the influence of the anatomical position on dispersion values, results are first averaged according to the position along the bone axis and then to the quadrant position for the three specimens. A diagram of the dispersion values as a function of the position along the bone axis (proximal, medial, or distal position) is shown in Fig. 4, which displays the results obtained in the axial direction (this direction is of importance in the context of axial transmission). The highest dispersion values are obtained in the distal part of the bone whereas the smallest dispersion values can be found in the centeroproximal part of the bone. Similar results are obtained for the two other directions of propagation (data not shown). Moreover, dispersion values (averaged according to the position along the bone axis) as a function of the quadrant position are shown in Fig. 5 for the axial direction. Dispersion values are the highest in the posterolateral position and the smallest in the anterior part. Again, similar results are obtained for the two other directions of propagation (data not shown). In Fig. 4 (respectively, Fig. 5), the error bars correspond to the standard deviation of the distribution of the dispersion values obtained for each anatomical location. All error bars in Figs. 4 and 5 are higher than the mean value of measurement uncertainties (see Table I). The ANOVA test reveals a significant anatomical position effect of the position around the bone circumference ($p<2\times10^{-6}$, $F=25$; $p<10^{-10}$, $F=45$; and $p<3\times10^{-4}$, $F=13$ for the axial, radial, and tangential directions, respectively) and for the position along the bone axis ($p<2\times10^{-2}$, $F=2.7$; $p<8\times10^{-3}$, $F=3.6$; and $p<3\times10^{-2}$, $F=2.6$ for the axial, radial, and tangential directions, respectively).

TABLE I. Average, standard deviation, and minimal and maximal dispersion values for the three measurement directions: axial, radial, and tangential. Dispersion is evaluated between 3.5 and 4.5 MHz.

	Dispersion ($m\,s^{-1}$)			
	Mean	SD	min	max
Axial	8	6.3	−3.3	45.5
Radial	6.9	5.3	0.7	25
Tangential	6.3	9.2	−21.1	42

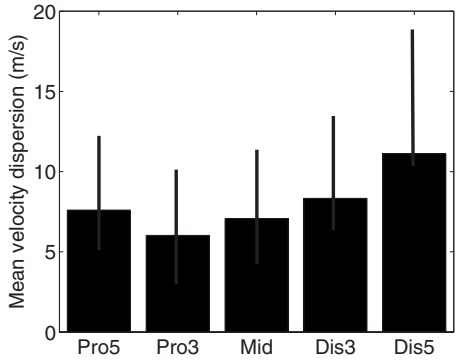


FIG. 4. Mean dispersion values and standard deviation (solid line on the bar diagram) as a function of the position along the bone axis for the axial direction. Pro5 and pro3 correspond to the cortical rings in the proximal part at 5 and 3 cm from the central part of the shaft. Mid corresponds to the cortical ring at the center part of the shaft. Dis5 and dis3 correspond to the cortical rings in the distal part at, respectively, 5 and 3 cm from the central part of the shaft.

The influence of bone microstructure on dispersion measurements is evaluated by averaging dispersion values for each histological group defined in Sec. II C for all samples. The average and standard deviation of dispersion values for each histological group are summarized in Table II, together with density and vBMD measurements. In Table II, the dispersion values are averaged for all anatomical positions. For the axial and radial directions of propagation, plexiform mi-

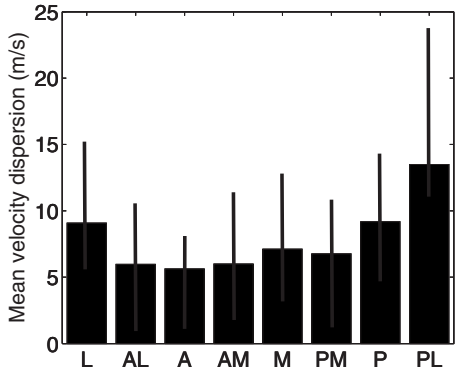


FIG. 5. Mean dispersion values and standard deviation (solid line on the bar diagram) as a function of the quadrant position for the axial direction. L is the lateral quadrant, AL is the anterolateral quadrant, A is the anterior quadrant, AM is the anteromedial quadrant, M is the medial quadrant, PM is the posteromedial quadrant, P is the posterior quadrant, and PL is the posterolateral quadrant.

TABLE II. Average and standard deviation of the dispersion values in the axial, radial, and tangential directions together with density and apparent volumetric bone mineral density measured by DXA for the four histological groups. The last column shows each quantity averaged for all samples.

	Plexiform	Haversian	Porotic	Mixed	All structures pooled
Dispersion ($m\ s^{-1}$)					
Axial	$6.1 \pm 3.9^{a,b}$	10.6 ± 8.3	14.7 ± 6.1^{c}	8.5 ± 5.8^{c}	8 ± 6.3
Radial	$5.6 \pm 4.6^{d,a}$	9.6 ± 6.0^{c}	$14.3 \pm 5.5^{c,b}$	5.8 ± 2.4^{a}	6.9 ± 5.3^{a}
Tangential	5.8 ± 6.3^{a}	5.6 ± 8.0	$20.7 \pm 13.9^{c,b}$	4 ± 9.9^{a}	6.3 ± 9.2^{a}
Density ($g\ cm^3$)	$2.09 \pm 0.02^{d,a}$	$2.03 \pm 0.04^{c,a}$	$1.95 \pm 0.05^{c,d}$	$2.06 \pm 0.02^{c,d,a}$	2.06 ± 0.05
BMD ($g\ cm^3$)	$1.47 \pm 0.04^{d,a}$	$1.42 \pm 0.06^{c,a}$	$1.29 \pm 0.06^{c,d}$	$1.47 \pm 0.06^{d,a}$	1.45 ± 0.07

[a]Significantly different from Po group.
[b]Significantly different from M group.
[c]Significantly different from Pl group.
[d]Significantly different from H group.

crostructures give the lowest mean dispersion values; Haversian dispersion values are greater than plexiform dispersion values but smaller than porotic dispersion. However, results in the tangential direction are slightly different as plexiform and Haversian dispersion values are similar and mixed microstructures give the lowest values of dispersion. vBMD and density variations are similar but evolve in the opposite way compared to dispersion values obtained for each histological group in the axial and radial directions. ANOVA analysis reveals a significant effect of microstructure on dispersion values ($p<3\times10^{-2}$, $F=2.9$; $p<10^{-6}$, $F=9.3$; and $p<2\times10^{-3}$, $F=5.7$ for the axial, radial, and tangential directions, respectively). As shown in Table II, for the axial direction, the Tuckey–Kramer analysis reveals that dispersion values are significantly different ($p<0.001$) for plexiform and porotic microstructures and for plexiform and mixed microstructures. No significant difference is found for the other microstructures. For the radial direction, dispersion values are significantly different ($p<10^{-5}$) for plexiform and porotic microstructures, for plexiform and Haversian microstructures, and for mixed and porotic microstructures. For the tangential direction, dispersion values are significantly different ($p<0.003$) for plexiform and porotic microstructures and for mixed and porotic microstructures.

Dispersion values are significantly correlated with vBMD and with mass density in the radial direction; the determination coefficient (R^2) between dispersion and vBMD in the radial direction is equal to 0.4 (root-mean-square error: RMSE=3.3 $m\ s^{-1}$, $p<10^{-5}$). However, dispersion is not correlated with vBMD in the axial and tangential directions since the determination coefficients between dispersion and vBMD are equal to 0.05 and 0.06, respectively. Dispersion values are plotted in Fig. 6 versus vBMD for the radial direction. Each histological group is indicated by a different marker and the dashed line shows the least squares linear regression line. The determination coefficient (R^2) between dispersion and mass density is equal to 0.33 (RMSE $=3.5\ m\ s^{-1}$, $p<10^{-5}$) in the radial directions.

IV. DISCUSSION

To the best of our knowledge, the present work represents the first study focusing on dispersion measurements in bovine cortical bone. It is also the first time that negative values of dispersion are measured in cortical bone.

The frequency bandwidth (3.5–4.5 MHz) used herein is similar to the one used previously in a study on the frequency dependence of the attenuation coefficient using the same bone samples [Sasso *et al.*, 2007; Sasso *et al.*, 2008). As discussed in our previous study (Sasso *et al.*, 2007), the value of the center frequency (4 MHz) was chosen to reach a compromise between a sufficiently small wavelength (around 1 mm) so that the ultrasonic wave is sensitive to bone heterogeneities and the requirement of an acceptable signal-to-noise ratio for all samples and all directions (always higher than 8 dB). Similarly, the frequency range (1 MHz) was a compromise between a satisfactory linear variation of the apparent phase velocity versus frequency for all samples (the determination coefficient is always higher than 0.96) and a sufficient amount of information contained within the bandwidth.

Many research groups focused on dispersion studies in trabecular bone (Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Droin *et al.*, 1998; Wear, 2000a; Wear, 2001b; Haïat *et al.*, 2006; Wear, 2007a), but considerably less attention has been given to cortical bone. As far as we know, dispersion is one of the less understood acoustical features of bone. A pioneering work was carried out by Yoon and Katz 1976 in which the authors were able to measure the ultrasonic velocity as a function of frequency between 2 and 10 MHz in

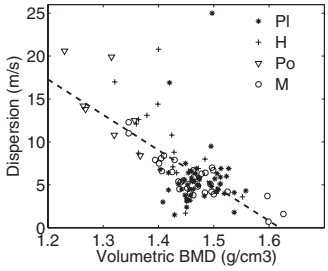


FIG. 6. Dependence of dispersion values as a function of bone mineral density values for each histological group in the radial direction. Stars correspond to the plexiform microstructure, crosses to the Haversian microstructure, triangles to the porotic microstructure, and circles to the mixed microstructure. The dashed line shows the linear regression analysis obtained with all data.

TABLE III. Characteristics of the samples where negative values of dispersion are measured with our ultrasonic device. The dispersion value and standard deviation and the direction of propagation are also indicated.

Bovine No.	Axial position	Quandrant position	Type of structure	Direction of propagation	Dispersion value ($m\ s^{-1}$)
1	Pro5	P	M	Tan	-13 ± 3.5
1	Dis5	A	M	Tan	-12.4 ± 5.9
1	Dis5	AM	M	Tan	-5.2 ± 2.7
1	Dis5	AL	M	Tan	-10.9 ± 3.2
2	Dis5	A	M	Tan	-7 ± 4
2	Dis5	AM	M	Ax	-3.3 ± 1.1
2	Dis5	AM	M	Tan	-5.5 ± 1.9
3	Dis5	AM	Ha	Tan	-21.1 ± 4.8
3	Dis5	PM	M	Tan	-16.2 ± 6.1

three samples of cortical bone. They found values of dispersion around 13 $m\ s^{-1}$ which spans within the range of the present study. Another work was carried out by Lakes *et al.* (1986) where the authors measured ultrasonic velocity in cortical bone between 1 and 12 MHz for about ten different frequencies. However, it is not easy to discuss these data in terms of dispersion due to the scale of the graph (see Fig. 2 of Lakes *et al.*, 1986). In a more recent study, the "sonic" velocity (Lees and Klopholz, 1992) was measured using different transducers in four samples of bovine cortical bone between 5 and 10 MHz. In the axial direction, dispersion between 5 and 10 MHz was found to be around $8 \pm 4.9\ m\ s^{-1}$, which is also in good agreement with our results. Their results in the radial direction ($13 \pm 6.2\ m\ s^{-1}$) are slightly higher but of the same order than the values found in the present study. However, no reports could be found in literature on negative values of dispersion in cortical bone.

A. Negative dispersion values

In this study, negative values of dispersion were obtained in nine samples. Table III shows the characteristics of the samples where negative dispersion is measured, together with the mean value and standard deviation (corresponding to the reproducibility) of the dispersion and the direction of propagation. The reproducibility of the measurement obtained for the samples where negative dispersion is measured is always higher than the average reproducibility. However, this relatively poor precision does not affect the main conclusion obtained on negative dispersion because in all cases, the standard deviation is lower in absolute value than the mean dispersion value. Negative values of dispersion are measured mostly in the tangential direction of propagation (eight samples out of nine). The resulting wider range of variation of dispersion values for the tangential direction explains the important standard deviation of the distribution of the dispersion values obtained in the tangential direction compared to the radial direction (see Tables I and II). Moreover, about 89% of the samples for which negative values of dispersion are measured are of mixed microstructure, which is higher than the total proportion of mixed microstructure (26%).

A possible explanation of the negative values of dispersion has been given in a recent study in the context of trabecular bone (Marutyan *et al.*, 2006). In their work, the authors nicely showed that negative values of dispersion may result from the interference of two broadband ultrasonic pulses arriving on the receiver with a given time delay. In their paper, both pulses were due to the presence of fast and slow wave modes propagating in trabecular bone. A similar interpretation could be applied to the case of cortical bone, which is a heterogeneous medium where wave splitting may occur, when different parts of the wave front propagate at different sound speeds (Yamato *et al.*, 2005; Yamato *et al.*, 2006). This effect is likely to be more pronounced for the mixed structures which are inherently more heterogeneous. In the tangential direction, the mean apparent phase velocities are equal to 3617 ± 113, 3513 ± 162, and $3283 \pm 147\ m\ s^{-1}$ in plexiform, Haversian, and porotic microstructures, respectively. Two ultrasonic pulses may interfere with a time delay depending on the respective velocity in both media and on the sample length, leading to a comparable situation as the one obtained by Marutyan *et al.* (2006). Figure 7 illustrates the overlapping of these two wave forms by showing an example of radio frequency (rf) signal corresponding to a dispersion value of $-19.1\ m\ s^{-1}$. Two broadband wave forms of different amplitudes seem to interfere. In the case of Fig. 7, the amplitude of the fast wave form is lower than the

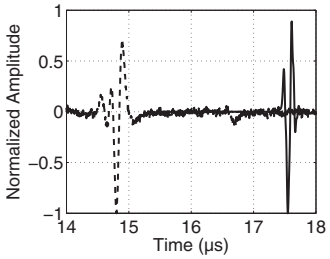


FIG. 7. Example of normalized rf signals obtained for a mixed microstructure. The black dashed line represents a wave form obtained through a bone sample in the tangential direction associated with a negative value of the dispersion ($-19.1\ m\ s^{-1}$). The ratio of the maximum amplitude of the signal transmitted through bone and through water is of 0.04. The sample was taken from the posteromedial part at 5 cm from the center on the diaphysis toward the distal end; the direction of propagation was tangential. The reference wave form is obtained through NSS (solid line).

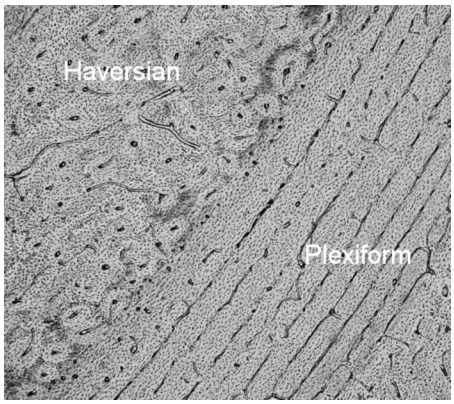


FIG. 8. Example of an optical microscopy obtained for a mixed microstructure. The image corresponds to a plane perpendicular to the axial direction. The size of the measured area is 3×4 mm^2.

amplitude of the slow wave form. Similarly as in the case of trabecular bone, a clear distinction between the two wave forms is not possible for all samples. However, the frequency components of both wave forms shown in Fig. 7 seem to be similar, whereas in cancellous bone, the nominal frequency of the fast wave has been shown to be lower compared to that of the slow wave (Hosokawa and Otani, 1997; Hosokawa and Otani, 1998; Haiat *et al.*, 2008b).

The structural organization of bovine cortical bone has been shown to be approximately axially symmetric (Locke, 2004). Figure 8 shows an optical microscopy corresponding to a mixed microstructure in a plane perpendicular to the axial direction. The model of Marutyan *et al.* (2006), in combination with the description of the microstructural organization of cortical bone, may explain (i) why most samples with negative dispersion are of mixed microstructure and (ii) why all bone samples with negative dispersion are measured in the axial or tangential direction. In addition, the relatively poor reproducibility obtained for samples with negative dispersion may also be roughly explained by the model of Marutyan *et al.* (2006), who showed that dispersion was strongly sensitive to the relative amplitude of both interfering components. The amplitude of the two wave fronts may be affected by a slightly different positioning of the sample relative to the transducer locations.

Here, negative values of dispersion were never measured in porous microstructure, which might seem surprising because porous microstructures are known to lead to the propagation of two longitudinal wave modes [as described by the Biot theory (Biot, 1956a; Biot, 1956b)], which may lead to negative dispersion (Marutyan *et al.*, 2006). However, the possibility of measuring two separated longitudinal wave modes propagating in trabecular bone depends on a complex combination of several factors, such as bone volume fraction, direction of propagation, and structural anisotropy (Haiat *et al.*, 2008b). There is no evidence to suggest that these conditions are fulfilled in the case of cortical bone in the framework of our experimental setup. In addition, to our knowledge, two longitudinal wave modes have not yet been observed in cortical bone.

Another possible physical explanation for negative values of dispersion may be multiple scattering phenomena,

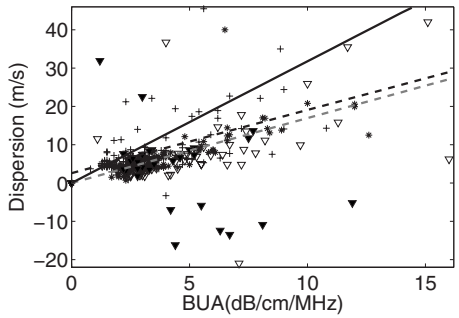


FIG. 9. Dependence of dispersion values as a function of BUA values for each direction of propagation. Crosses correspond to the axial direction of propagation and stars to the radial one. Filled triangles correspond to samples of mixed microstructure measured in the tangential direction and open triangles correspond to samples of other microstructure measured in the tangential direction. The solid black line shows the results obtained by using the Kramers–Kronig relationships. The dashed black and gray lines show the linear regression between dispersion and BUA values obtained, respectively, in the radial ($R^2=0.73$, $p<10^{-5}$) and axial ($R^2=0.28$, $p<10^{-5}$) directions (no correlation was obtained in the tangential direction).

which are known to be responsible for negative values of dispersion in trabecular bone (Haïat *et al.*, 2007a). However, further investigation is needed to evidence such effect as no reports could be found in literature on multiple scattering in cortical bone.

B. Relationship with broadband ultrasonic attenuation

Broadband ultrasonic attenuation (BUA) measurements (given by the slope of the frequency-dependent attenuation coefficient versus frequency) have been previously performed on the same samples (Sasso *et al.*, 2007; Sasso *et al.*, 2008) and are compared with dispersion results obtained herein. Figure 9 shows the relationship between dispersion and BUA values for all bone samples and in the three directions of propagation. Crosses correspond to the axial direction of propagation, stars to the radial direction, and triangles to the tangential one (full and hollow triangles, respectively, indicate mixed microstructure and other microstructures).

1. Correlation BUA/dispersion

The dashed black and gray lines of Fig. 9 show the linear regression between dispersion and BUA values obtained, respectively, in the radial ($R^2=0.73$, $p<10^{-5}$) and axial ($R^2=0.28$, $p<10^{-5}$) directions. The linear regression obtained for the tangential direction is not indicated because no correlation was found for this direction.

2. Relation between BUA, dispersion, and microstructure

In a previous study (Sasso *et al.*, 2008), BUA has been shown to be significantly correlated with vBMD for the three directions (R^2 comprised between 0.55 and 0.67). The correlation between BUA and vBMD, together with the correlation between dispersion and BUA in the radial direction shown in the present study, may explain our results, showing that the dispersion and vBMD values are correlated in the radial direction.

BUA values measured in plexiform structures have been shown to be significantly lower than in the other microstructures (Sasso *et al.*, 2008). Similarly, the (positive) correlation between BUA and dispersion in the axial and radial directions may explain why dispersion in plexiform structures is also lower than in other microstructures (see Table II). The absence of correlation between BUA and dispersion in the tangential direction explains the different variations of dispersion as a function of microstructure obtained in this direction (see Table II).

3. Using Kramers–Kronig relationship?

In order to better understand our results, we used the formulas obtained in O'Donnell *et al.*, 1978 and O'Donnell *et al.*, 1981, where the authors used the local Kramers–Kronig (KK) relationships to derive an expression between the frequency dependence of the attenuation coefficient and that of phase velocity, assuming (i) that the attenuation coefficient is known over the entire frequency bandwidth, (ii) a linear system, (iii) that causality is respected, (iv) that the attenuation and dispersion do not vary rapidly as a function of frequency, and (v) an homogenous material. Assuming a linear variation of the attenuation coefficient versus frequency over the entire bandwidth, the KK relationships lead to a logarithmic variation of phase velocity as a function of frequency. The slope D of the phase velocity at 4 MHz as a function of frequency is an estimate of dispersion and is then given by (O'Donnell *et al.*, 1978; O'Donnell *et al.*, 1981)

$$D = \frac{100 \times \mathrm{BUA} \times V_b^2(4\ \mathrm{MHz})}{8.68\pi^2 f}, \tag{6}$$

provided that BUA and f are, respectively, given in dB MHz^{-1} cm^{-1} and in MHz. The solid black line in Fig. 9 corresponds to the representation of the relation given by Eq. (6). Here, V_b (4 MHz) is chosen equal to 3305 m s^{-1}, because this value corresponds approximately to the mean phase velocity for the radial direction, which exhibits the best correlation between BUA and dispersion.

The KK relationships predict an increase in dispersion values when BUA increases, which is consistent with the results shown in Fig. 9 for the radial and axial directions. However, the slope of the linear fit obtained between the dispersion and BUA values is overestimated by the KK relationships compared to the experimental results in the axial and radial directions.

Several reasons may explain this discrepancy. First, KK relationships cannot be applied when two waves overlap (as it is the case shown in Fig. 7), resulting in phase cancellation effects. Bone is intrinsically a tremendously heterogeneous structure in which the Kramers–Kronig argument must be applied with caution. Second, the nonlinear variation of the attenuation coefficient (Sasso *et al.*, 2007) as a function of frequency over the entire frequency bandwidth, which is an assumption used to derive Eq. (6), may also cause this same discrepancy. More generally, it seems that some difficulties may come from the interpretation of the general Kramers–Kronig relationships using a linear frequency dependence for attenuation and dispersion. The linear fit of phase velocity in the considered frequency range is actually an approximation of the real dispersion. In addition, the signal spectra used here are inherently bandwidth limited by the transducers and therefore it was necessary to extend the attenuation law from the measured bandwidth over all frequencies (Waters *et al.*, 2000).

V. CONCLUSION

This *in vitro* study shows the feasibility of measurements of velocity dispersion in cortical bone samples obtained from three specimens. Measurements were performed in a 1 MHz wide bandwidth centered on 4 MHz. Dispersion is shown to depend significantly on the anatomical location. Mostly positive dispersion values are obtained. However, negative dispersions were also measured, mostly in samples constituted of mixed microstructure in the tangential direction of propagation. The dependence of dispersion on the broadband ultrasonic attenuation was compared with results derived from the local Kramers–Kronig relationships. The results presented in the present study might help understand the ultrasonic propagation in this complex (heterogenous and viscoelastic) medium, which may lead to new development in the domain of ultrasonic devices applied to bone characterization.

ACKNOWLEDGMENTS

Part of this study was supported by the Academic Frontier Research Project on "New Frontier of Biomedical Engineering Research" of Doshisha University and by the Agence Nationale de la Recherche (Contract BoneChar No. BLAN06-2_144779). Mami Matsukawa would like to thank the JSPS and CNRS bilateral joint project and JSPS Grant in Aid for Scientific Research (B) for financial support. Magali Sasso would like to thank the University Paris 12—Val de Marne ("Conseil Scientifique") for support during her stay at the Laboratory of Ultrasonic Electronics, Doshisha University, Japan.

Ashman, R. B., Cowin, S. C., Van Buskirk, W. C., and Rice, J. C. (**1984**), "A continous wave technique for the measurement of the elastic properties of cortical bone," J. Biomech. **17**, 349–361.

Barkmann, R., Kantorovich, E., Singal, C., Hans, D., Genant, H. K., Heller, M., and Gluer, C. C. (**2000**), "A new method for quantitative ultrasound measurements at multiple skeletal sites: first results of precision and fracture discrimination," J Clin Densitom **3**, 1–7.

Bensamoun, S., Gherbezza, J.-M., de Belleval, J.-F., and Ho Ba Tho, M.-C. (**2004a**), "Transmission scanning acoustic imaging of human cortical bone and relation with the microstructure," Clin. Biomech. (Bristol, Avon) **19**, 639–647.

Bensamoun, S., Ho Ba Tho, M.-C., Luu, S., Gherbezza, J.-M., and de Belleval, J.-F. (**2004b**), "Spatial distribution of acoustic and elastic properties of human femoral cortical bone," J. Biomech. **37**, 503–510.

Biot, M. A. (**1956a**), "Theory of propagation of elastic waves in a fluid-satured porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178.

Biot, M. A. (**1956b**), "Theory of propagation of elastic waves in a fluid-satured porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191.

Bossy, E., Talmant, M., Defontaine, M., Patat, F., and Laugier, P. (**2004a**), "Bidirectional axial transmission can improve accuracy and precision of ultrasonic velocity measurement in cortical bone: a validation on test materials," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 71–79.

Bossy, E., Talmant, M., Peyrin, F., Akrout, L., Cloetens, P., and Laugier, P.

(**2004b**), "An in vitro study of the ultrasonic axial transmission technique at the radius: 1-MHz velocity measurements are sensitive to both mineralization and intracortical porosity," J. Bone Miner. Res. **19**, 1548–1556.
Cadossi, R., de Terlizzi, F., Cane, V., Fini, M., and Wuster, C. (**2000**), "Assessment of bone architecture with ultrasonometry: experimental and clinical experience," Horm. Res. **54**, 9–18.
Cowin, S. C. (**2001**), *Bone Mechanics Handbook* (CRC, Boca Raton, FL).
Cullum, I. D., Ell, P. J., and Ryder, J. P. (**1989**), "X-ray dual-photon absorptiometry: a new method for the measurement of bone density," Br. J. Radiol. **62**, 587–592.
Currey, J. (**1984**), "Comparative mechanical properties and histology of bone," Am. Zool. **24**, 5–12.
Droin, P., Berger, G., and Laugier, P. (**1998**), "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592.
Duck, F. A. (**1990**), *Physical Properties of Tissue* (University Press, Cambridge, UK).
Foldes, A. J., Rimon, A., Keinan, D. D., and Popovtzer, M. M. (**1995**), "Quantitative ultrasound of the tibia: A novel approach for assessment of bone status," Bone (N.Y.) **17**, 363–367.
Glüer, C. C., Blake, G., Lu, Y., Blunt, B. A., Jergas, M., and Genant, H. K. (**1995**), "Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques," Osteoporosis Int. **5**, 262–270.
Haiat, G., Padilla, F., Barkmann, R., Kolta, S., Latremouille, C., Gler, C. C., and Laugier, P. (**2005**), "In vitro speed of sound measurement at intact human femur specimens," Ultrasound Med. Biol. **31**, 987–996.
Haïat, G., Padilla, F., Cleveland, R., and Laugier, P. (**2006**), "Effect of frequency dependent attenuation and dispersion on different speed of sound measurements on human intact femur," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 39–51.
Haïat, G., Padilla, F., Laugier, P., Lonné, S., and Naili, S. (**2007a**), "Modeling the frequency dependence of phase velocity in phantoms of trabecular bone," International Congress on Ultrasonics, Vienna, Austria.
Haiat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2007b**), "Variation of ultrasonic parameters with microstructure and material properties of trabecular bone: a 3D model simulation," J. Bone Miner. Res. **22**, 665–674.
Haïat, G., Padilla, F., Peyrin, F., and Laugier, P. (**2008b**), "Fast wave ultrasonic propagation in trabecular bone: numerical study of the influence of porosity and structural anisotropy," J. Acoust. Soc. Am. **123**, 1694–1705.
Hosokawa, A., and Otani, T. (**1997**), "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562.
Hosokawa, A., and Otani, T. (**1998**), "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722.
Kanis, J. A. (**2002**), "Diagnosis of osteoporosis and assessment of fracture risk," Lancet **359**, 1929–1936.
Kaufman, J. J., Xu, W., Chiabrera, A. E., and Siffert, R. S. (**1995**), "Diffraction effects in insertion mode estimation of ultrasonic group velocity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 232–242.
Kaye, G. W. C., and Laby, T. H. (**1973**), *Table of Physical and Chemical Constants* (Longmans, London, UK).
Kino, G. S. (**1987**), *Acoustic Waves: Devices, Imaging, and Analog Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Lakes, R. S., Yoon, H. S., and Katz, J. L. (**1986**), "Ultrasonic wave propagation and attenuation in wet bone," J. Biomed. Eng. **8**, 143–148.
Langton, C. M., Palmer, S. B., and Porter, R. W. (**1984**), "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.
Laugier, P., Berger, G., Giat, P., Bonnin-Fayet, P., and Laval-Jeantet, M. (**1994**), "Ultrasound attenuation imaging in the os calcis: an improved method," Ultrason. Imaging **16**, 65–76.
Lee, S. C., Coan, B. S., and Bouxsien, M. L. (**1997**), "Tibial ultrasound velocity measured in situ predicts the material properties of tibial cortical bone," Bone (N.Y.) **21**, 119–125.
Lees, S., Ahern, J. M., and Leonard, M. (**1983**), "Parameters influencing the sonic velocity in compact calcified tissues of various species," J. Acoust. Soc. Am. **74**, 28–33.
Lees, S., Cleary, P. F., Heeley, J. D., and Gariepy, E. L. (**1979**), "Distribution of sonic plesio-velocity in a compact bone sample," J. Acoust. Soc. Am. **66**, 641–646.
Lees, S., and Klopholz, D. Z. (**1992**), "Sonic velocity and attenuation in wet compact cow femur for the frequency range 5 to 100 MHz," Ultrasound Med. Biol. **18**, 303–308.
Lefebvre, F., Deblock, Y., Campistron, P., Ahite, D., and Fabre, J. J. (**2002**), "Development of a new ultrasonic technique for bone and biomaterials in vitro characterization," J. Biomed. Mater. Res. **63**, 441–446.
Locke, M. (**2004**), "Structure of long bones in mammals," J. Morphol. **262**, 546–565.
Marutyan, K. R., Holland, M. R., and Miller, J. G. (**2006**), "Anomalous negative dispersion in bone can result from the interference of fast and slow waves," J. Acoust. Soc. Am. **120**, EL55–61.
Moilanen, P., Nicholson, P. H., Kilappa, V., Cheng, S., and Timonen, J. (**2006**), "Measuring guided waves in long bones: modeling and experiments in free and immersed plates," Ultrasound Med. Biol. **32**, 709–719.
Morse, P., and Ingard, K. (**1986**), *Theoretical Acoustics* (Princetown University Press, Pricetown, NJ).
Nakamura, Y., and Otani, T. (**1993**), "Study on the surface elastic wave induced on backing and diffracted field of a piezoelectric polymer film hydrophone," J. Acoust. Soc. Am. **94**, 1191–1199.
Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and Van der Perre, G. (**1996**), "A comparison of time-domain and frequency domain approaches to ultrasonic velocity measurement in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.
Nicholson, P. H. F., and Njeh, C. F. (**1999**), "Ultrasonic studies of cancellous bone in vitro," in *Quantitative Ultrasound: Assesment of Osteoporosis and Bone Status*, edited by C. F. Njeh, D. Hans, T. Fuerst, C. C. Gluer, and H. K. Genant (Martin Dunitz Ltd., London), pp. 195–220.
Njeh, C. F., Hans, D., Fuerst, T., and Genant, H. K. (**1999**), *Quantitative Ultrasound: Assessment of Osteoporosis and Bone Status* (Martin Dunitz Ltd., London).
O'Donnell, M., Jaynes, E. T., and Miller, J. G. (**1978**), "General relationships between ultrasonic attenuation and dispersion," J. Acoust. Soc. Am. **63**, 1935–1937.
O'Donnell, M., Jaynes, E. T., and Miller, J. G. (**1981**), "Kramers–Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701.
Protopappas, V. C., Fotiadis, D. I., and Malizos, K. N. (**2006**), "Guided ultrasound wave propagation in intact and healing long bones," Ultrasound Med. Biol. **32**, 693–708.
Rico, H. (**1997**), "The therapy of osteoporosis and the importance of cortical bone," Calcif. Tissue Int. **61**, 431–432.
Sasso, M., Haïat, G., Yamato, Y., Naili, S., and Matsukawa, M. (**2007**), "Frequency dependence of ultrasonic attenuation in bovine cortical bone: an in vitro study," Ultrasound Med. Biol. **33**, 1933–1942.
Sasso, M., Haïat, G., Yamato, Y., Naili, S., and Matsukawa, M. (**2008**), "Dependence of ultrasonic attenuation on bone mass and microstructure in bovine cortical bone," J. Biomech. **41**, 347–355.
Strelitzki, R., and Evans, J. A. (**1996**), "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.
Ta, D. A., Huang, K., Wang, W. Q., Wang, Y. Y., and Le, L. H. (**2006**), "Identification and analysis of multimode guided waves in tibia cortical bone," Ultrasonics **44**, e279–e284.
Verhoef, W. A., Cloosterman, J. T. M., and Thijssen, J. M. (**1985**), "Diffraction and velocity dispersion effects on the estimation of ultrasound attenuation and velocity in biological tissues," IEEE Trans. Biomed. Eng. **32**, 521–529.
Waters, K. R., Hughes, M. S., Brandeburger, G. H., and Miller, J. G. (**2000**), "On the applicability of Kramers–Krönig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563.
Wear, K. (**2000a**), "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.
Wear, K. A. (**2000b**), "The effects of frequency-dependent attenuation and dispersion on sound speed measurements: applications in human trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 265–273.
Wear, K. A. (**2001a**), "A numerical method to predict the effects of frequency-dependent attenuation and dispersion on speed of sound estimates in cancellous bone," J. Acoust. Soc. Am. **109**, 1213–1218.
Wear, K. A. (**2001b**), "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083.
Wear, K. A. (**2007a**), "Group velocity, phase velocity, and dispersion in human calcaneus in vivo," J. Acoust. Soc. Am. **121**, 2431–2437.
Wear, K. A. (**2007b**), "The dependence of time-domain speed-of-sound measurements on center frequency, bandwidth, and transit-time marker in human calcaneus in vitro," J. Acoust. Soc. Am. **122**, 636–644.

Xu, W., and Kaufman, J. J. (**1993**), "Diffraction correction methods for insertion ultrasound attenuation estimation," IEEE Trans. Biomed. Eng. **40**, 563–570.
Yamato, Y., Kataoka, H., Matsukawa, M., Yamazaki, K., Otani, T., and Nagano, A. (**2005**), "Distribution of longitudinal wave velocities in bovine cortical bone in vitro," Jpn. J. Appl. Phys., Part 1 **44**, 4622–4624.
Yamato, Y., Matsukawa, M., Otani, T., Yamazaki, K., and Nagano, A. (**2006**), "Distribution of longitudinal wave properties in bovine cortical bone in vitro," Ultrasonics **44**, 233–237.
Yamato, Y., Matsukawa, M., Yanagitani, T., Yamazaki, K., and Nagano, A. (**2008**), "Distribution of hydroxyapatite crystallites orientation and ultrasonic wave velocity ring-shaped cortical bone of bovine femur," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 1298–1303.
Yoon, H. S., and Katz, L. (**1976**), "Dispersion of the ultrasonic velocities in human cortical bone," 1976 Ultrasonic Symposium Proceeding IEEE, pp. 48–50.

Whistle sequences in wild killer whales (*Orcinus orca*)

Rüdiger Riesch
Department of Zoology, University of Oklahoma, 730 Van Vleet Oval, Norman, Oklahoma 73019

John K. B. Ford
Cetacean Research Program, Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia V9T 6N7, Canada

Frank Thomsen[a)]
Center for Environment, Fisheries and Aquaculture Science (Cefas), Pakefield Road, Lowestoft, Suffolk, NR33 0HT, United Kingdom

(Received 1 February 2008; revised 30 May 2008; accepted 9 June 2008)

Combining different stereotyped vocal signals into specific sequences increases the range of information that can be transferred between individuals. The temporal emission pattern and the behavioral context of vocal sequences have been described in detail for a variety of birds and mammals. Yet, in cetaceans, the study of vocal sequences is just in its infancy. Here, we provide a detailed analysis of sequences of stereotyped whistles in killer whales off Vancouver Island, British Columbia. A total of 1140 whistle transitions in 192 whistle sequences recorded from resident killer whales were analyzed using common spectrographic analysis techniques. In addition to the stereotyped whistles described by Riesch *et al.*, [(2006). "Stability and group specificity of stereotyped whistles in resident killer whales, *Orcinus orca*, off British Columbia," Anim. Behav. **71**, 79–91.] We found a new and rare stereotyped whistle (*W*7) as well as two whistle elements, which are closely linked to whistle sequences: (1) *stammers* and (2) *bridge* elements. Furthermore, the frequency of occurrence of 12 different stereotyped whistle types within the sequences was not randomly distributed and the transition patterns between whistles were also nonrandom. Finally, whistle sequences were closely tied to close-range behavioral interactions (in particular among males). Hence, we conclude that whistle sequences in wild killer whales are complex signal series and propose that they are most likely emitted by single individuals.
 [DOI: 10.1121/1.2956467]

PACS number(s): 43.80.Jz, 43.80.Ka [WWA] Pages: 1822–1829

I. INTRODUCTION

Combining different stereotyped vocal signals into specific sequences increases the range of information that can be transferred between individuals. With regard to message possibilities, using established signals is even more efficient than producing new ones (Hauser, 1997; Bradbury and Vehrencamp, 1998). However, in order to ascribe a specific function to vocal sequences, at least two prerequisites have to be fulfilled: (1) the signals within the sequence have to follow a specific and nonrandom pattern, and (2) the behavioral context in which the sequence takes place has to be identified.

Both the temporal emission pattern and the behavioral context of vocal-signal sequences have been described in detail for a variety of birds and mammals, where they are often called "songs" and serve a variety of functions (general reviews by Hauser, 1997; Bradbury and Vehrencamp, 1998; songbirds: Catchpole and Slater, 1995; Slater, 2003; terrestrial mammals: Marler and Tenaza, 1977; Byrne, 1982; Geissmann and Orgeldinger, 2000; Gourbal *et al.*, 2004; Holy and Guo, 2005; marine mammals: Tyack, 1998).

For cetaceans, the study of vocal sequences is still in its infancy: they have been described for some species such as bowhead whales (*Balaena mysticetus*), fin whales (*Balaenoptera physalus*), bottlenose dolphins (*Tursiops truncatus*), killer whales (*Orcinus orca*), and humpback whales (*Megaptera novaeangliae*) (Lilly and Miller, 1961; Payne and McVay, 1971; Bain, 1986; Ford, 1989; Tyack, 1998; Miller *et al.*, 2004); yet, only for the latter two were more detailed insights into their function provided. In killer whales, vocal sequences are often comprised of repetitions of similar stereotyped calls by different members within a social group and are probably used to coordinate group movements (Ford, 1989; Miller *et al.*, 2004). Best studied are the songs of male humpback whales by which females gain information about the location of the singer, as well as his willingness to breed and compete with other males for females (Tyack, 1998).

Resident killer whales off British Columbia produce burst-pulsed calls and tonal sounds called whistles in underwater communications (Ford, 1989). Pulsed calls can be either variable or stereotyped (discrete) in structure. Discrete calls can be quite intense (>160 dB re: 1 μPa at 1 m; Miller, 2006) and are proportionally most often used during behaviors where animals are widely spaced out such as traveling and foraging. Repertoires of stereotyped pulsed calls are

[a)]Electronic mail: frank.thomsen@cefas.co.uk

group specific and probably function as long-range contact signals and in group affiliation (Ford, 1989, 1991; Miller *et al.*, 2004). Whistles on the other hand are high pitched, complex, and comparably low-intensity sounds that are almost entirely associated with close-range interactions among individuals (Ford, 1989; Thomsen *et al.*, 2001, 2002; Miller, 2006). Recently, we demonstrated the presence of 15 stereotyped whistle types in resident killer whales off Vancouver Island, with some types unchanged in spectrographic contour over a period of at least 15 years (Riesch *et al.*, 2006). We also showed that killer whales that have no overlap in their call repertoire use essentially the same set of stereotyped whistles. Based on these results, we suggested that stereotyped whistles provide a community-level means of recognition that facilitates association and affiliation of members of different clans, which otherwise use different communicative signals (Riesch *et al.*, 2006). In the process of data analysis, we constantly noted repetitive and rather elaborate series of stereotyped whistles that occurred during socializing (Ford, 1991; Riesch *et al.*, 2006). However, a detailed analysis of whistle sequences was lacking to date. Therefore, the function of whistle sequences for underwater communication in wild killer whales was completely unknown.

In the present study, we provide a detailed analysis of whistle sequences in resident killer whales off Vancouver Island, British Columbia. We examine the frequency of occurrence of different whistle types and the transition patterns of stereotyped whistles within the sequences. We also consider the behavioral context in which whistle sequences were produced. Based on our results, we discuss the likely function of whistle sequences in underwater communication of resident killer whales.

II. METHODS

A. Acoustic data collection

Our study animals belong to the northern community of resident killer whales, which ranges from mid-Vancouver Island north to southeastern Alaska and consisted at the time of the study of 216 individuals in 33 matrilines (1999 census; Ford *et al.*, 2000). The majority (70%–95%) of burst-pulsed calls is highly stereotyped and can be assigned to different call types, called "discrete calls" (Ford and Fisher, 1983; Ford, 1989). Based on discrete pulsed call similarities, the northern resident killer whales are grouped into three distinct vocal clans: A-clan is by far the largest, followed by G-clan, and the smallest is the R-clan (Ford, 1991).

Fieldwork was undertaken in western Johnstone Strait and adjacent waters, British Columbia (50° 30′ N, 126° 35′ W) in 1996–1997 and in 2001 and 2003. A total of 281 field trips were conducted with more than 1300 h spent at sea. Killer whales were observed on 254 field excursions with a total of >250 h observation time. Killer whale individuals were identified by visual inspection of natural markings on the dorsal fin and back (Ford *et al.*, 2000). Underwater sounds were recorded using digital and analog recorders (1996: Sony TCD-D8, 1997: Sony TCD-D7, 2001/2003: Sony WMD-6C) and three different hydrophones (1996: Bruel & Kjaer 8101, 1997/2003: Offshore Acoustics, 2001: DEEPSEA Powerlight hydrophone (SM 1000 S/N 153); sensitivities: (−180 dB re: 1 V/μPa or greater; frequency responses: 20 Hz–18/20 kHz ± 1 dB). Simultaneous voice recordings of behavioral observations were made on a separate track of the same tape (for more details see Thomsen *et al.*, 2002; Riesch *et al.*, 2006). In addition, we used recordings obtained between 1978–1983 and 1993–1999 by one of us (J.K.B.F.). V. Deecke generously provided additional tapes of northern residents obtained between 1999 and 2001. Details of recording equipment and methodology are given by Ford (1989, 1991) and Deecke *et al.* (2000, 2005)

B. Acoustic analyses

More than 90 h of killer whale recordings were initially scanned for whistle sequences using real-time spectrographic analysis (Raven 1.2, Cornell Laboratory of Ornithology; sample rate=50 kHz, frequency range=0–22 kHz, dynamic range=42 dB, FFT size=512 points; window type =Hanning). Based on previous analysis (e.g., Thomsen *et al.*, 2001, 2002; Riesch *et al.*, 2006), we defined *a priori* that a sequence had to consist of at least two whistles occurring within 5.0 s of each other. Structural parameters of 192 whistle sequences in 41 recordings from 1978 to 2003 that had a good signal-to-noise ratio were further analyzed. Whistles were classified according to their spectrographic contour, and defined as being either stereotyped, variable, *stammers*, or *bridge* elements (please refer to Sec. III for definitions of stammers and bridge elements). Stereotyped whistles were repetitive with a distinct spectrographic contour. These were classified alphanumerically as *W*1 (whistle-type 1), *W*2, and so on (Riesch *et al.*, 2006). Furthermore, some whistle types occur in two versions, either with or without a trill-like ending. Hence, some stereotyped whistles are denoted with a *T* to indicate the version with a trill at the end (e.g., *W*1 exists as *W*1 or $W1_T$; Riesch *et al.*, 2006).

For each sequence, the duration of each whistle and each interval between whistles was measured. Also, the numbers of stereotyped whistles, variable whistles, stammers, and bridge elements were counted for each sequence. Furthermore, the numbers of pure whistle transitions (defined as the transition between two stereotyped whistles), mixed whistle transitions (defined as the transition between a stereotyped whistle and a nonstereotyped whistle), and impure whistle transitions (defined as the transition between two nonstereotyped whistles) were counted.

C. Temporal emission patterns within the sequence

For a total of 1140 whistle transitions, the observed numbers of transitions of each whistle type to itself and all other whistle types were compared to a random distribution using an χ^2-test. However, only *W*1, *W*3, $W3_T$, bridge elements, stammers, and variable whistles could be compared in this way, because all other whistle types had expected values below 5 (Quinn and Keough, 2002). Furthermore, the degree of uncertainty in predicting what whistle type follows after a given whistle type was estimated by information theory procedures (H_2-statistic; see Frick and Miller, 1951; Attneave, 1959; Devenport and Merriman, 1983). The H_2-statistic

scores range between 0 and 1, with low scores implying rigid sequences and 1 implying complete randomness/uncertainty (Frick and Miller, 1951; Attneave, 1959; Devenport and Merriman, 1983). This statistic is sensitive to the degree to which one vocal signal is predictive of the following vocal signal. In a first step, a first-order uncertainty value (H_1) is calculated that describes whether the first vocal signal in a sequence is chosen randomly or based on some underlying pattern (as described above, values range from 0 to 1). In a second step, the second-order uncertainty value (H_2) is calculated that describes if the first signal in a two-signal sequence has an impact on the identity of the second signal. Finally, a χ^2-analysis is applied that tests whether the total sequence differs significantly from chance, or in other words if the predictability of the second signal increases with knowledge of the first (Frick and Miller, 1951; Attneave, 1959; Devenport and Merriman, 1983). However, other studies used the H_2-statistic on binary choices, whereas in the present study there were 15 different possibilities; therefore, the analyses were carried out using $\log_{15}$ instead of $\log_2$.

Furthermore, we ran an additional χ^2-analysis that tested whether the transitions from an individual given whistle type to the subsequent whistle type deviated significantly from chance. Since the H_2-statistic only provides information whether there is a general pattern in the transitions, this test was designed to identify if certain whistle types follow a more rigid pattern of transition than others. However, to meet the χ^2-assumptions that not more than 20% of the expected values were below 5 (Quinn and Keough, 2002), the six least common whistle types ($W1_T$, $W2_T$, $W4$, $W5$, $W6_T$, and $W7$) were grouped into one category for this analysis.

D. Whistle sequences and activity state

The behavioral activities of the northern resident killer whales were grouped into six categories: beach rubbing, foraging, resting, socializing, social traveling, and traveling (defined after Ford, 1989; Barrett-Lennard *et al.*, 1996; Thomsen *et al.*, 2002). Activity states were recorded *ad libitum* (Martin and Bateson, 1993) whenever a general change in group activity was observed. In all cases, the activity state of the recorded group of whales was considered to be the activity of most group members. Thus, for a subset of 489 recordings, a distinct behavioral state could be ascribed, while 46 recordings had to be removed from the analysis due to lack of information on the whales' activity state.

The observed frequency of behavioral activities was compared to a random distribution across behavioral states for a total of 489 recordings, which were extracted from more than 90 h of *ad libitum* recording of killer whale vocal behaviors. The observed and expected frequencies of behavioral states were then compared using a χ^2-test.

III. RESULTS

A. Identification of new whistle types

In addition to the stereotyped whistle types described in (Riesch *et al.*, 2006), we found a rare new stereotyped whistle, which was called *W7* (Fig. 1). *W7* was found only ten times in six different recordings in the whole data set; however, only five *W7*s were strong enough in their signal-to-noise ratio for acquiring measurements ($N=5$, "start frequency"$=5270.6\pm324.6$ Hz (mean ± S.D.), "end frequency"$=5098.4\pm288.2$ Hz, "minimum frequency"$=4900.4\pm123.1$ Hz, "maximum frequency"$=6700.6\pm436.8$ Hz, "bandwidth"$=1800.2\pm376.1$ Hz, "carrier frequency"$=7106.0\pm4587.0$ Hz, "duration"$=757.4\pm164.2$ ms, "frequency modulations" $=4.6\pm0.9$; Fig. 1). However, there was no particular affiliation of this whistle type with specific killer whale groups (at least five different matrilines emitted this whistle) or behaviors (it occurred during traveling, social traveling, and socializing).

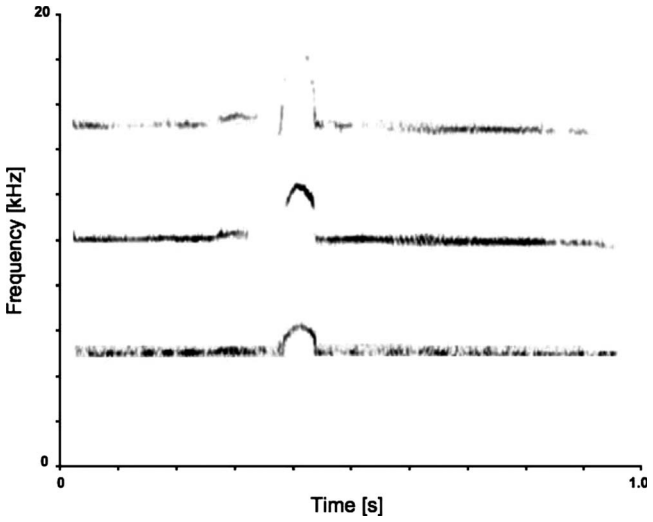


FIG. 1. Representative spectrogram of the new stereotyped whistle-type *W7* (frequency resolution=98 Hz, time resolution=10.2 ms, FFT size=512 points, window type=Hanning).

Furthermore, we found whistles that only occurred as part of whistle sequences: (1) stammers and (2) bridge elements (Fig. 2). Both are not stereotyped whistles in the common sense, but are categories that we defined for this analysis. Stammers look like the beginning of a *W1*, $W1_T$, *W3*, or $W3_T$ whistle; however, the characteristic downsweep and/or trill-like ending is never produced (stammers: $N=140$, "whistle duration" $=0.97\pm0.25$ s; Fig. 2; Riesch *et al.*, 2006). Often, stammers can be found at the start or end of a sequence. Bridge elements, on the other hand, seem to be used as connecting pieces that link individual stereotyped whistles within a sequence (bridge elements: $N=173$, whistle duration$=0.82\pm0.30$ s; Fig. 2). Even though bridge elements seem highly variable in contour they were grouped together for this analysis.

B. Parameters of whistle sequences

The vast majority (84%) of all stereotyped whistles found in the recordings were associated with whistle sequences, while only 16% of stereotyped whistles appeared as isolated whistles. Figure 2 shows representative spectrograms of whistle sequences. The 192 whistle sequences had an average duration of 9.84 ± 7.43 s (mean ± S.D.) and consisted of 6.16 ± 3.44 whistle elements (3.91 ± 1.99 stereotyped whistles, 1.37 ± 1.15 variable whistles, 1.35 ± 1.20 stammers, and 1.10 ± 0.85 bridge elements), with an average

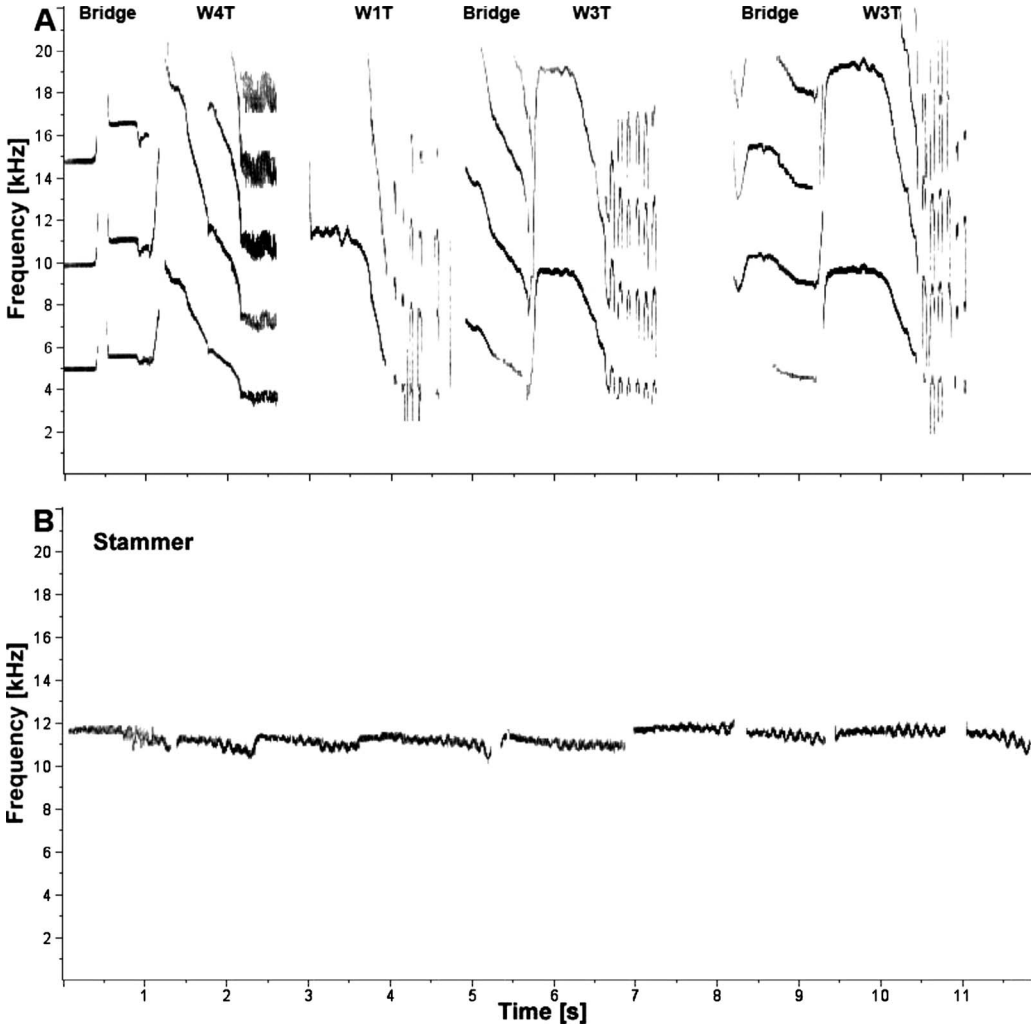


FIG. 2. Representative spectrograms of whistle sequences (frequency resolution=98 Hz, time resolution=10.2 ms, FFT size=512 points, window type =Hanning). A: sequence of bridge-$W4_T$-$W1_T$-brdge-$W3_T$-bridge-$W3_T$; B: sequence of stammers.

of 2.57 ± 0.83 different stereotyped whistle types per sequence. Whistle sequences comprised 2.09 ± 1.42 pure transitions, 2.73 ± 2.04 mixed transitions, and 1.31 ± 1.23 impure transitions. The duration of the intermission between two whistles within a sequence was 0.36 ± 0.21 s and on average 23.63 ± 14.51 animals were present during recordings that contained whistle sequences.

C. Frequency of occurrence of whistle types

The frequency of occurrence of different stereotyped whistle types within the sequences was not randomly distributed ($\chi^2_{11}=657.094$, $N=1321$, $P<0.001$; Fig. 3). Some whistle types were rather scarce ($W4$, $W5$, and $W7$), while others were predominant ($W1$, $W3$, $W3_T$, and $W4_T$; Fig. 3).

D. Temporal emission patterns within the sequence

For six whistle types ($W1$, $W3$, $W3_T$, bridge elements, stammers, and variable whistles), the transition patterns to following whistles differed significantly from the expected random distribution (Tables IA and IB). The transition patterns of the remaining whistle types ($W1_T$, $W2$, $W4_T$ and the grouped whistle types: $W2_T$, $W4$, $W5$, $W6_T$, and $W7$) did not differ significantly from randomness (Table IA). For certain pairs of whistle types, the difference between observed and expected frequency was greater than for other pairs (Table

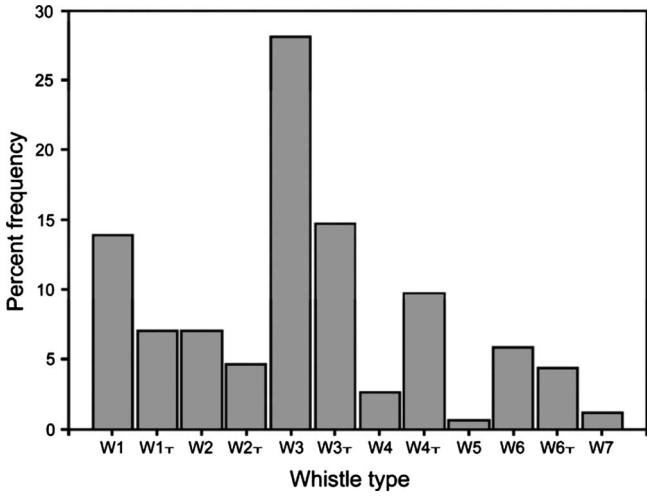


FIG. 3. Frequency of occurrence of stereotyped whistle types during 192 whistle sequences.

TABLE I. Contingency table analysis of transitions among whistle types within 192 whistle sequences. Transition frequency matrix for 1140 whistle transitions. A: total transitions. B: listing of only the significant Chi-square results for the comparison between the expected and observed transition patterns.

A Preceeding whistle	Following whistle														
	*W*1	$W1_T$	*W*2	$W2_T$	*W*3	$W3_T$	*W*4	$W4_T$	*W*5	*W*6	$W6_T$	*W*7	Bridge	Stammer	Variable
*W*1	7	7	3	3	17	14	1	3	1	7	0	1	28	8	7
$W1_T$	7	5	4	1	1	2	1	0	0	2	1	0	6	5	7
*W*2	3	1	3	3	11	7	0	2	0	3	0	1	3	5	3
$W2_T$	4	4	2	1	5	2	0	0	0	1	0	1	2	3	3
*W*3	10	3	10	8	45	19	0	11	0	5	2	1	57	15	22
$W3_T$	3	5	12	6	8	12	0	1	0	2	2	1	17	8	15
*W*4	1	0	1	0	1	1	4	3	0	2	2	0	1	1	2
$W4_T$	6	4	5	3	9	3	0	9	0	3	3	4	7	1	6
*W*5	0	0	0	0	0	0	1	2	0	0	1	0	1	0	0
*W*6	6	1	3	3	0	1	0	3	0	2	16	0	6	0	4
$W6_T$	2	2	0	0	3	0	3	4	0	1	1	0	1	0	4
*W*7	2	0	3	1	0	0	0	0	0	0	0	0	2	0	2
Bridge	16	7	2	2	62	25	2	26	1	5	3	0	13	5	2
Stammer	18	2	1	1	20	4	0	0	0	2	0	0	4	81	5
Variable	8	5	3	5	23	15	3	7	2	5	2	0	6	4	58

B Preceeding whistle	Chi-square test			
	N	χ^2	*df*	*P*
*W*1	107	111.60	14	<0.001
*W*3	208	280.01	14	<0.001
$W3_T$	92	72.71	14	<0.001
Bridge	171	328.56	14	<0.001
Stammer	138	661.13	14	<0.001
Variable	146	303.03	14	<0.001

IA). Stammers, for example, were almost exclusively found at the beginning and at the end of sequences while others, such as bridge elements, predominantly connected *W*3's, $W3_T$'s, and $W4_T$'s with each other within the sequences. Furthermore, the majority of variable whistles were successive, and some pairs of whistles were highly stereotyped in themselves, as is the case with the *W*6-$W6_T$-pair.

On the first order of estimation the uncertainty (H_1) was 87%, while on the second order the uncertainty (H_2) was only 72%. However, the second-order probability was significantly different from the first-order probability (χ^2_{81} =756.461, N=1140, $P<0.001$).

E. Whistle sequences and activity state

The occurrence of the 192 whistle sequences within the 41 recordings was closely linked to certain activity states (Fig. 4). Contrary to the expected random distribution, whistle sequences occurred predominantly during socializing and social traveling, but only rarely during foraging and traveling (χ^2_5=35.554, N=41, $p<0.001$; see also Thomsen *et al.*, 2002). We were not able to identify the whistle-emitting individuals for the recordings analyzed for this study. However, in some cases we were able to appoint a certain group of whales as the most likely source of the recorded whistle sequences (based on changes in the sound intensity of whistle sequences in combination with movement patterns of certain subgroups of whales). Interestingly, in all six (15% overall) of these cases, the emitting group of whales was an all-male group. Furthermore, in all cases males were present and in at least four other cases, an all-male group was present during the recording [6 (all-male groups only) +4 (all-male groups present) +31 (at least one male present)=41 recordings].

IV. DISCUSSION AND CONCLUSIONS

This study describes a new stereotyped whistle (*W*7) in northern resident killer whales, and two relative broad categories of stereotyped whistles produced as part of whistle sequences (stammers and bridge elements). However, since

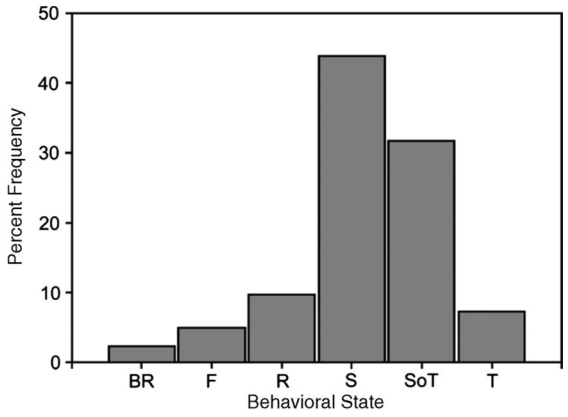


FIG. 4. Relative occurrence of whistle sequences during different activity states (N=41 recordings). BR: beach rubbing; F: foraging; R: resting; S: socializing; SoT: social traveling; T: traveling.

we could not find an association of them with particular behaviors or usage by particular groups, no further conclusions on possible functions of these signals can be drawn yet. We could further show that sequences of stereotyped whistles consist of whistle types that follow a nonrandom pattern of emission. Finally, we found some indications that sequences are particularly associated with male-male social behaviors within socializing and social traveling behavioral states.

Whistle sequences in killer whales are rather complex: Parts of the sequences are made up of multiloops, which are repetitions of the same whistle type (Table IA). Multilooped stammers are predominantly emitted at the start or end of a sequence (personal observation). Hence, they might serve as a "lead-in" and "lead-out" for the killer whales in such a way that they open and close the acoustical channel used for the transmission of whistle sequences. Another form of multiloops is composed of *W*4's or *W*3-bridge-, $W3_T$-bridge-, and $W4_T$-bridge-combinations, respectively. These different multiloops might be used to enhance and stress the respective information encoded in the whistle types within the sequences. However, this does not mean that sequences merely consist of different multilooped whistles put together. On the contrary, whistle sequences are quite complex and consist of a variety of different whistle combinations, some of which are rather common (e.g., $W6$-$W6_T$; Table IA) while others are either rare or not found at all (e.g., *stammer*-*W*4; Table IA).

There are three possible explanations for how whistle sequences might be formed. First, several members within a group could contribute to them. Thus, whistle sequences would be structurally similar to the choruses found in several terrestrial mammals where they serve to synchronize behaviors (Estes and Goddard, 1967; Mitani and Gros-Louis, 1998). Although several studies already suggest that in resident killer whales discrete calls serve in behavioral synchronization especially during long-range communication (Ford, 1989, 1991; Miller *et al.*, 2004; Miller, 2006), whistle sequences may have this function during close-range interactions. However, if whistle sequences represent choruses, we would expect to find overlapping whistles within them, which was almost never the case. Finally, the energy levels of whistles within a given sequence were more or less the same throughout (Fig. 2). If sequences represent choruses, we would expect to find whistles of different intensities within a sequence, since a variety of animals of different sizes and different positions to the hydrophone (resulting in different sound levels) would "join in." This was clearly not the case. Thus, even though the whistle sequences could be used to synchronize behaviors, we do not think that they represent choruses.

Second, sequences could be formed by two individuals. If so, then two interacting individuals could be answering each other with the same or with other whistle types. Such whistling behavior has been described for common dolphins (*Delphinus delphis*; Caldwell and Caldwell, 1968) and bottlenose dolphins (*Tursiops truncatus*; Lilly and Miller, 1961; Janik, 2000). In gibbons and songbirds, duets are used for pair bonding ceremonies, agonistic interactions, territorial defense, and mate defense against rivals of the same sex (Estes and Goddard, 1967; Marler and Tenaza, 1977; Arrowood, 1988; Malacarne *et al.*, 1991; Geissmann and Orgeldinger, 2000; Slater, 2003). However, the results of this study do not seem to support this interpretation since only a few whistle types had the tendency to follow one another more often than different types. In a different scenario, two individuals could answer each other's whistles with a different whistle type rather than the same. If that were the case, whistle sequences could potentially be formed by two or more individuals. However, even though we do not have any direct evidence for or against this possibility, we think that this scenario is rather unlikely.

We propose therefore that whistle sequences are mainly emitted by single individuals and may at the most be answered by another sequence from a different whale, which is close by. In particular, the nonexistent overlap of different whistles and the continuous energy level within the sequences (see Fig. 2) support this possibility.

Compared to the loud songs of humpback whales (Au *et al.*, 2006), gibbons (e.g., McAngus Todd and Merker, 2004), and songbirds (e.g., Brackenbury, 1979), killer whale whistle sequences are relatively low in sound pressure levels (Thomsen *et al.*, 2001; Miller 2006). Thus they may be used as more intimate signals between a limited number of close individuals, which would vastly limit the number of eavesdroppers listening in on the sequences (for a review on public and private signals refer to Dabelsteen, 2005). If this were the case, then these sequences would almost definitely not be used as advertisement displays such as songs in other species. It is possible that whistles are body-contact enjoyment sounds, such as laughter in chimpanzees (*Pan troglodytes*: Marler and Tenaza, 1977; Goodall, 1986) or purring in cats (Peters, 1984); however, the elaborate nature of whistle sequences and the stereotypy of a variety of different whistle signals do not support this idea.

In this context, it should be noted that all members within one community probably share the same set of stereotyped whistles (Riesch *et al.*, 2006). Thus, information provided by them might be potentially available to all members of the community and it is unclear then why such a rather universal signal should be kept private. It is noteworthy that sequences were often heard when groups of interacting males were in the vicinity, and often were the only group within range of the hydrophone. These male-only social interactions are quite frequent in northern resident killer whales and usually involve at least one adult male and one or more adolescents [age class assigned according to age-related morphological changes: Olesiuk *et al.*, 1990; for details on all-male groups see Rose, N.A. (1992). Ph.D. thesis, University of California, Santa Cruz, CA (unpublished)]. Rose (1992) proposed that male-only social interactions serve an affiliative function, since agonistic behavior was not observed during surface interactions and males from all age groups (primarily unrelated individuals) were involved. However, one cannot rule out that these are competitive but highly ritualized interactions where, for example, some of the males might compete for access to breeding females. Whistle sequences might function as a means to coordinate these interactions. Here, they could encode the affiliative or agonistic/competitive motivation of the signaler with long and complex sequences

representing a higher motivation than shorter and simpler ones. Communication during these events may be kept private to prevent or minimize attraction by rival males. Alternatively, whistle sequences might function to initiate or strengthen male-specific relationships, and to give adolescents the possibility to learn male-specific cognitive and social skills (Rose, 1992). Finally, if these represent alliances of males used to herd females for mating, as described for bottlenose dolphins (e.g., Connor *et al.*, 1992), male talk should also be kept rather quiet so as not to alert potential competitors to a mating site.

A more detailed look at bridge elements may lead them to being grouped into several structurally similar categories, as structural variability was strong. Finally, future studies using underwater cameras together with hydrophone arrays could help us clarify the specific function of whistle sequences in their repertoire of social behavior among individual males.

ACKNOWLEDGMENTS

We would like to thank Bill and Donna Mackay and Jim Borrowman for their support for this study. Thanks also go to Robert Butler, Wayne Garton, Rolf Hicker, Tyson Mackay, Brian Sylvester, Dave Tyre, and Steve Wischniowski for their help in the field. Dave Briggs, Anna Spong, Paul Spong, Helena Symonds, and Rob Williams provided important information on whale locations. Rüdiger Riesch would like to thank Lynn D. Devenport, Martin Plath, Ingo Schlupp, and Michael Tobler for their help during the analysis at the University of Oklahoma. Volker Deecke supplied additional sound recordings from 1999 to 2001. This study was partly funded by scholarships from the "Deutscher Akademischer Austauschdienst" (DAAD), the University of Hamburg (Graduiertenstipendium), and the "Hansische Universitätsstiftung."

Arrowood, P. C. (**1988**). "Duetting, pair bonding and agonistic display in parakeet pairs," Behaviour **106**, 129–157.

Attneave, F. (**1959**). *Application of Information Theory to Psychology* (Rine, Rinehart and Winston, New York).

Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M. H., and Andrews, K. (**2006**). "Acoustic properties of humpback whale song," J. Acoust. Soc. Am. **120**, 1103–1110.

Bain, D. E. (**1986**). "Acoustic behavior of *Orcinus*: Sequences, periodicity, behavioral correlates and an automated technique for call classification," in *Behavioral Biology of Killer Whales*, edited by B. C. Kirkevold and J. S. Lockard (Alan R. Liss, Inc., New York), pp. 335–371.

Barrett-Lennard, L. G., Ford, J. K. B., and Heise, K. A. (**1996**). "The mixed blessing of echolocation: Differences in sonar use by fish-eating and mammal-eating killer whales," Anim. Behav. **51**, 553–565.

Brackenbury, J. H. (**1979**). "Power capabilities of the avian sound-producing system," J. Exp. Biol. **78**, 163–166.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer Associates, Sunderland, MA).

Byrne, R. W. (**1982**). "Primate vocalisations: Structural and functional approaches to understanding," Behaviour **80**, 241–258.

Caldwell, M. C., and Caldwell, D. K. (**1968**). "Vocalizations of naive captive dolphins in small groups," Science **159**, 1121–1123.

Catchpole, C. K., and Slater, P. J. B. (**1995**). *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge).

Connor, R. C., Smolker, R. A., and Richards, A. F. (**1992**). "Two levels of alliance formation among male bottlenose dolphins (*Tursiops* sp.)," Proc. Natl. Acad. Sci. U.S.A. **89**, 987–990.

Dabelsteen, T. (**2005**). "Public, private or anonymous? Facilitating and countering eavesdropping," in *Animal Communication Networks*, edited by P. K. McGregor (Cambridge University Press, Cambridge), pp. 38–62.

Deecke, V. B., Ford, J. K. B., and Spong, P. (**2000**). "Dialect change in resident killer whales: Implications for vocal learning and cultural transmission," Anim. Behav. **60**, 629–638.

Deecke, V. B., Ford, J. K. B., and Slater, P. J. B. (**2005**). "The vocal behaviour of mammal-eating killer whales: Communicating with costly calls," Anim. Behav. **69**, 395–405.

Devenport, L. D., and Merriman, V. J. (**1983**). "Ethanol and behavioral variability in the radial-arm maze," Psychopharmacology (Berlin) **79**, 21–24.

Estes, R. D., and Goddard, J. (**1967**). "Prey selection and hunting behavior of the african wild dog," J. Wildl. Manage. **31**, 52–70.

Ford, J. K. B. (**1989**). "Acoustic behaviour 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.

Ford, J. K. B., Ellis, G. M., and Balcomb III, K. C. (**2000**). *Killer Whales*, 2nd ed. (UBC Press, Vancouver).

Ford, J. K. B., and Fisher, H. D. (**1983**). "Group-specific dialects of killer whales (*Orcinus orca*) in British Columbia," in *Communication and Behavior of Whales*, edited by R. Payne (West View Press Inc., Boulder, Colorado), pp. 129–161.

Frick, F. C., and Miller, G. A. (**1951**). "A statistical description of operant conditioning," Am. J. Psychol. **64**, 20–36.

Geissmann, T., and Orgeldinger, M. (**2000**). "The relationship between duet songs and pair bonds in siamangs, *Hylobates syndactylus*," Anim. Behav. **60**, 805–809.

Goodall, J. (**1986**). *The Chimpanzees of Gombe: Patterns of Behavior* (The Belknap Press of Harvard University Press, Cambridge).

Gourbal, B. E. F., Barthelemy, M., Petit, G., and Gabrion, C. (**2004**). "Spectrographic analysis of the ultrasonic vocalisations of adult male and female BALB/c mice," Naturwiss. **91**, 381–385.

Hauser, M. D. (**1997**). *The Evolution of Communication* (MIT Press, Cambridge, MA).

Holy, T. E., and Guo, Z. (**2005**). "Ultrasonic songs of male mice," PLoS Biol. **3**, 2177–2186.

Janik, V. M. (**2000**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.

Lilly, J. C., and Miller, A. M. (**1961**). "Vocal exchanges between dolphins," Science **134**, 1873–1876.

Malacarne, G., Cucco, M., and Camanni, S. (**1991**). "Coordinated visual displays and vocal duetting in different ecological situations among Western Palearctic non-passerine birds," Ethol. Ecol. Evol. **3**, 207–219.

Marler, P., Tenaza, R. (**1977**). "Signaling behavior of apes with special reference to vocalization," in *How Animals Communicate*, edited by T. A. Sebeok (Indiana University Press, Bloomington, IN), pp. 965–1033.

Martin, P., and Bateson, P. (**1993**). *Measuring Behavior: An Introductory Guide*, 2nd ed. (Cambridge University Press, Cambridge).

McAngus Todd, N. P., and Merker, B. (**2004**). "Siamang gibbons exceed the saccular threshold: Intensity of the song of *Hylobates syndactylus*," J. Acoust. Soc. Am. **115**, 3077–3080.

Miller, P. J. O. (**2006**). "Diversity in sound pressure levels and estimated active space of resident killer whale vocalizations," J. Comp. Physiol., A **192**, 449–459.

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.

Mitani, J. C., and Gros-Louis, J. (**1998**). "Chorusing and call convergence in chimpanzees: Tests of three hypotheses," Behaviour **135**, 1041–1064.

Olesiuk, P. F., Bigg, M. A., and Ellis, G. M. (**1990**). "Life history and population dynamics of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State," Rep. Int. Whal. Comm. **12**, 209–243.

Payne, R. S., and McVay, S. (**1971**). "Songs of the humpback whales," Science **173**, 585–597.

Peters, G. (**1984**). "On the structure of friendly close range vocalizations in terrestrial carnivores (Mammalia: Carnivora: Fissipedia)," Z. Säugetierkunde **49**, 157–182.

Quinn, G. P., and Keough, M. J. (**2002**). *Experimental Design and Data Analysis for Biologists* (Cambridge University Press, Cambridge).

Riesch, R., Ford, J. K. B., and Thomsen, F. (**2006**). "Stability and group specificity of stereotyped whistles in resident killer whales, *Orcinus orca*,

off British Columbia," Anim. Behav. **71**, 79–91.
Slater, P. J. B. (**2003**). "Fifty years of bird song research: A case study in animal behaviour," Anim. Behav. **65**, 633–639.
Thomsen, F., Franck, D., and Ford, J. K. B. (**2001**). "Characteristics of whistles from the acoustic repertoire of resident killer whales (*Orcinus orca*) off Vancouver Island, British Columbia," J. Acoust. Soc. Am. **109**, 1240–1246.
Thomsen, F., Franck, D., and Ford, J. K. B. (**2002**). "On the communicative significance of whistles in wild killer whales (*Orcinus orca*)," Naturwiss. **89**, 404–407.
Tyack, P. L. (**1998**). "Acoustic communication under the sea," in *Animal Acoustic Communication—Sound Analysis and Research Methods*, edited by S. L. Hopp, M. J. Owren, and C. S. Evans (Springer-Verlag, Berlin), pp. 163–220.

Automatic recognition of harmonic bird sounds using a frequency track extraction algorithm

Jason R. Heller
Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794-3600

John D. Pinezich
Advanced Acoustic Concepts, Inc., Hauppauge, New York 11788

(Received 1 May 2007; revised 29 May 2008; accepted 30 May 2008)

This paper demonstrates automatic recognition of vocalizations of four common bird species (herring gull [*Larus argentatus*], blue jay [*Cyanocitta cristata*], Canada goose [*Branta canadensis*], and American crow [*Corvus brachyrhynchos*]) using an algorithm that extracts frequency track sets using track properties of *importance and harmonic correlation*. The main result is that a complex harmonic vocalization is rendered into a set of related tracks that is easily applied to statistical models of the actual bird vocalizations. For each vocalization type, a statistical model of the vocalization was created by transforming the training set frequency tracks into *feature vectors*. The extraction algorithm extracts sets of frequency tracks from test recordings that closely approximate harmonic sounds in the file being processed. Each extracted set in its final form is then compared with the statistical models generated during the training phase using Mahalanobis distance functions. If it matches one of the models closely, the recognizer declares the set an occurrence of the corresponding vocalization. The method was evaluated against a test set containing vocalizations of both the 4 target species and 16 additional species as well as background noise containing planes, cars, and various natural sounds. [DOI: 10.1121/1.2950085]

PACS number(s): 43.80.Ka [MCH] Pages: 1830–1837

I. INTRODUCTION

An important application of automated bird sound recognition is as an aid to mitigating bird-aircraft strike hazards. There are more than 10 000 reported collisions of aircraft with birds every year costing more than $100 million in damage and canceled or delayed flights.[1] The US Air Force has implemented a program called Avian Hazard Advisory System (AHAS) that uses next generation weather radar to monitor the movements of large flocks of birds. An important part of AHAS is the bird avoidance model that shows bird concentrations for a specific time in a specific geographic region as an aid for planning flight paths.[2] Such models generally rely on radar and visual data for tracking and species identification. Little work has been done with acoustic methods for either tracking or identification; the research presented here was part of an Air Force program to explore such methods. Tracking of birds was accomplished using a microphone array design based on a passive acoustic underwater array used by the US Navy. Classification methods were developed based on Shamma's method of ripple analysis[3] and the spectrogram track extraction method discussed here. This paper describes a new algorithm for extracting harmonically related tracks from a spectrogram.

Some previous work on automated bird classification by vocalization has been done by Kogan and Margoliash,[4] Härmä and Somervuo,[5] Fagerlund,[6] and Chen and Maher.[7] Kogan and Margoliash used hidden Markov models (HMMs) to identify specific "syllables" within recordings of zebra finch (*Taeniopygia guttata*) and indigo bunting (*Passerina cyanea*), while Anderson, *et al.*[8] implemented dynamic time warping on these species. Shamma's neurologically inspired sound recognition algorithms have achieved good recognition results for the vocalizations of several bird species, including the species studied in this research. Härmä and Fagerlund have applied a method that uses frequency track based features to recognize syllables within a large database of Finnish songbird recordings. Chen and Maher's work also used frequency track based features.

Other bioacoustic pattern recognition methods include Buck and Tyack's method[9] of recognizing individual dolphins by their "signature whistles;" they extracted the curve defining the whistle from each whistle spectrogram and used that curve to classify the individual dolphin. Mellinger and Clark's method[10] of recognizing bowhead whale vocalizations in arbitrary recordings used a specialized form of spectrogram correlation (spectrogram correlation compares a template spectrogram to the spectrogram of the sound to be classified). Shamma's method of sound classification calculates a special neurologically inspired spectrogram, which is further processed and compared with a model generated during the training phase. Feature vectors are widely used in pattern recognition. For example, in HMM-based speech recognition,[11] feature vectors based on the mel-frequency cepstrum are used for classifying individual speech phonemes. Fristrup and Watkins' work[12] on marine mammal sound classification used a feature vector set specially devised for that work, as does this work.

TABLE I. SNRs for training set.

Species	SNRs (min,med,max) (dB)
American crow	7.31, 14.47, 28.33
Blue jay	7.14, 12.17, 16.11
Canada goose	7.16, 14.35, 26.85
Herring gull	16.11, 27.50, 30.11

II. THE DATA

A. Data collection

Recordings of American crow, blue jay, herring gull, Canada goose, and other species were obtained by the authors in Stony Brook, Northport, Sunken Meadow State Park, and at Heckscher State Park (all on Long Island, NY). The equipment used consisted of a Creative Nomad Jukebox storage unit, a Sound Devices preamplifier, and a Sennheiser directional shotgun microphone. All recordings were made in 22 050 Hz, 16 bit wave format. The preamplifier level was set to ensure that there was a good signal-to-noise ratio (SNR), but no saturation (except for rare occasions when a bird vocalized loudly nearby). Benson of Texas A&M University also provided a set of recordings of many different bird species (recorded using a parabolic dish microphone in 48 000 Hz, 16 bit .wav format). These were used to test the false alarm rate.

For each of the four species to be recognized, the recordings of that species were divided into two sets: one set for training the algorithm and the other for testing the trained algorithm. Both training and test recordings are processed using frequency track extraction and segment calculation as described below. Track sets from recordings for training are further processed by hand into species models as described in the following section. Track sets from recordings for testing are further processed automatically as described later. The SNRs for both the vocalizations of the training set and the vocalizations recognized from the test set were computed by setting the noise level to be the minimum rms voltage level of any 0.1 s segment in the whole recording. The start and stop time of each vocalization was set to be the start and stop time of the longest track in the vocalization. Table I lists the SNRs (min, median, and max) for the training set and Table II lists them for the test set.

B. Frequency track extraction

All data were preprocessed using a frequency track extraction (FTE) algorithm. The FTE algorithm transforms a time series into a spectrogram from which is extracted a set of *frequency tracks* in the time-frequency domain. Each track T is a set of 3-tuples of the form

$$T=\{(t_i,f_i,A_i)|i=0,1,\ldots,M-1\}, \quad (1)$$

where t stands for time, f stands for frequency, A stands for amplitude [discrete Fourier transform (DFT) magnitude], and $t_i<t_{i+1}$ for all i. The FTE is computed as follows.

Step 1. Calculate a DFT of the time series current frame. For more precise peak location, $N(v-1)$ zeros were appended to the end of each frame, where N is the length of the frame and v is an integer parameter ($v=8$ was the value used in all computations). For side lobe reduction, a Hanning window was used.

Step 2. In the current frame, calculate peaks from the local maxima in the DFT magnitude spectrum. The spectrum produced contains artifacts from noise, causing multiple peaks close together in frequency. These were reduced by smoothing the DFT magnitude spectrum using least squares piecewise cubic curve fitting (as described below) over a sliding window of 30 frequency points, resulting in a set of local maxima for each time t_i.

Step 3. Connect peaks in the current frame B to those in the previous frame A. Denote the set of peaks in frame A by $\{(t_A,f_1,a_1),(t_A,f_2,a_2),\ldots,(t_A,f_m,a_m)\}$ and those in frame B by $\{(t_B,g_1,b_1),(t_B,g_2,b_2),\ldots,(t_B,g_n,b_n)\}$, where f_i,g_i are frequencies and a_i,b_i are corresponding amplitudes. To determine which peak from frame B (if any) should be connected to (t_A,f_i,a_i) from frame A, find the nearest neighbor of (t_B,f_i,a_i) (in frequency) in frame B; call it (t_B,g_j,b_j). If the nearest neighbor to (t_B,g_j,b_j) in frame A is (t_A,f_i,a_i), then the two peaks should be connected. Otherwise, the track through (t_A,f_i,a_i) ends at (t_A,f_i,a_i). This particular algorithm for connecting peaks from consecutive frames was used by McAulay and Quatieri[13] for speech analysis and a slightly modified version of this algorithm was used by Serra and Smith[14] for analyzing the sounds produced by musical instruments.

Repeat steps 1–3 until there are no more frames left. Note that this process cannot connect tracks that cross, as could arise from two birds simultaneously vocalizing or two syringes within the same bird. The method is sufficient for the purposes of this paper, which deals with single vocalizations with noncrossing tracks.

Step 4. Remove short and weak tracks in a "pruning" process: all tracks shorter than a user-specified duration δ_t and all but a user-specified percentage μ of the strongest tracks are removed. Track T_1 is defined to be stronger than track T_2 if the maximum amplitude in track T_1 is greater than that in track T_2. Figures 1(a)–1(c) show the three steps of preprocessing: spectrogram, raw track extraction, and automatic track pruning using $\delta_t=0.1$ s and $\mu=100\%$ (in general, $\mu=50\%$ was used iteratively to reduce labor of computing training sets). For recognition, $\delta_t=0.04$ and $\mu=100\%$ were used.

TABLE II. SNRs for test set.

Species	SNRs (min,med,max) (dB)
American crow	3.13, 16.55, 26.05
Blue jay	4.54, 11.51, 23.44
Canada goose	2.84, 12.02, 20.92
Herring gull	15.73, 30.35, 32.80

C. Segment calculation

A track can be further approximated by a continuous piecewise linear function composed of line segments of al-

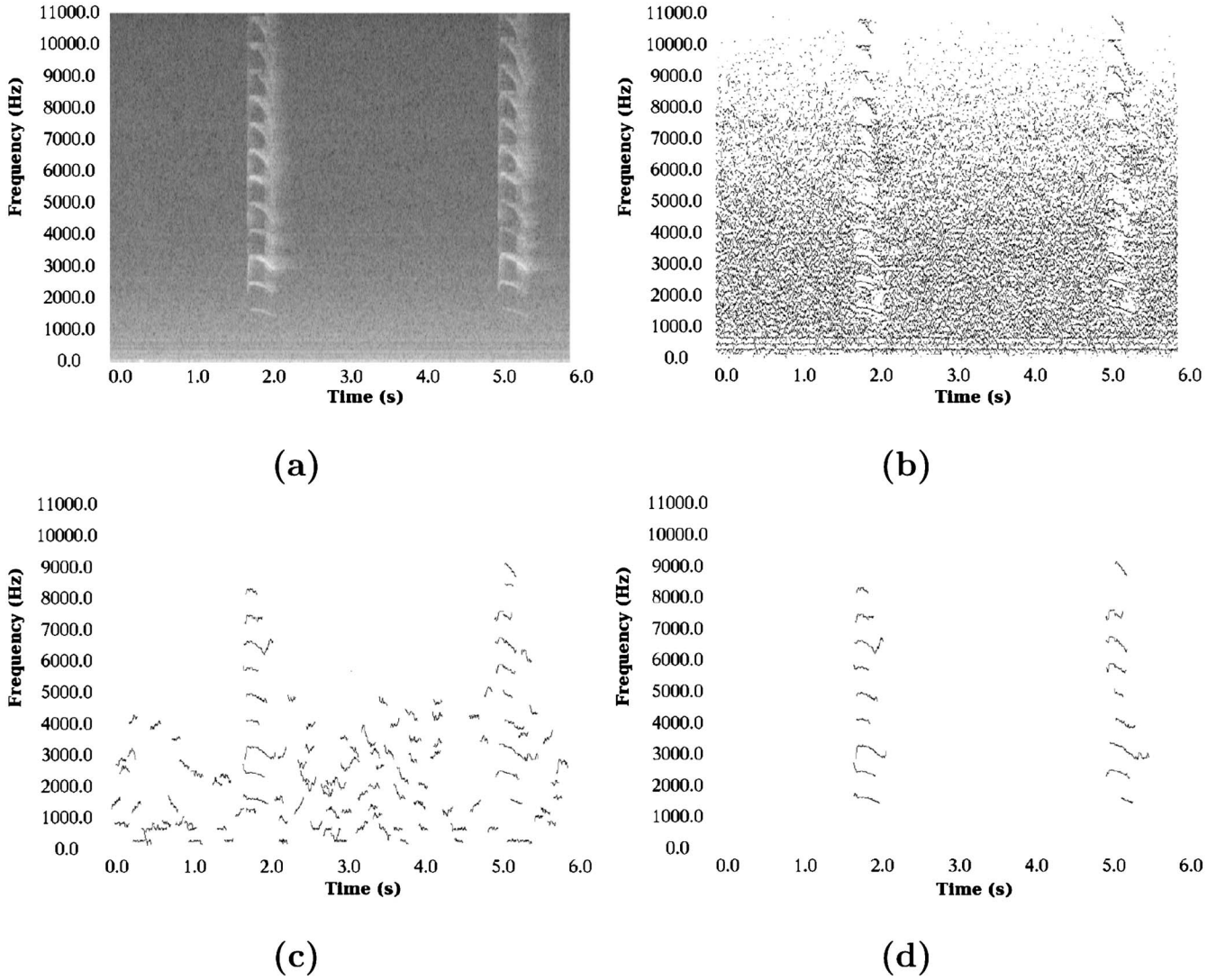


FIG. 1. Preprocessing and hand pruning for training, blue jay example: (a) spectrogram (step 1), (b) after generation (step 3), and (c) after pruning (step 4). (d) Hand pruning in training.

ternating slope signs. It was necessary to first eliminate short segments that were induced by noise; the remaining linear pieces are called *track segments*. The following algorithm was used with duration threshold $\beta_t=0.02$ s and frequency threshold $\beta_f=20$ Hz.

Step 1. Smooth the track using least squares piecewise cubic curve fitting with a window size of 20 spectrogram frames and a step size of 5 frames. The resulting curve fit is continuously differentiable because the initial slope of the cubic curve in each window is constrained to equal the final slope of the cubic curve in the previous window. The curve fit in the first window is a line so that there are two degrees of freedom for the curve fit in each window.

Step 2. From the saved cubic polynomial parameter sets for each track, calculate the local extrema (t_k,f_k) over the track.

Step 3. Mark extrema (t_{k+1},f_{k+1}) for deletion if $|t_{k+1}-t_k|<\beta_t$ and $|f_{k+1}-f_k|<\beta_f$. Endpoint extrema are not removed.

Step 4. After deleting and renumbering, further mark extrema (t_k,f_k) for deletion if f_k-f_{k-1} and $f_{k+1}-f_k$ have the same sign.

Step 5. Form a piecewise linear approximation of the track by joining the remaining neighboring extrema with straight lines, retaining durations $\tau_k=t_{k+1}-t_k$ and slopes $\delta_k=(f_{k+1}-f_k)/\tau_k$.

III. TRAINING

A. Harmonic structure

Training is an abstraction of the information captured in the spectrograms for each vocalization type, and results primarily in a statistical model for each harmonic. Assume K instances of vocalization type V are selected as a training set. For each instance i of V, a track set $\mathcal{T}_i$ is derived from step 4 above. The K track sets $\mathcal{T}_i$ are analyzed by hand to identify H harmonics particular to V, including possibly a track corresponding to the harmonic fundamental frequency. Once identified, each $\mathcal{T}_i$ is further pruned by hand to eliminate all tracks not consistent with the harmonics for V, resulting in

track sets $\bar{\mathcal{T}}_i$ containing $n_i \leqslant H$ valid tracks [see Fig. 1(d)]. The model for V is created from these K track sets.

For the four species of interest, the harmonic structure consisted of harmonics 0–9 for blue jay, 0–5 for herring gull, 0–7 for Canada goose, and 0 for American crow.

B. Descriptor sequence

Certain birds have highly repetitive segment slope signs in the spectrograms of their vocalizations, such as the herring gull and the Canada goose. For these species, a visual analysis is done to identify the *descriptor sequence*: an alternating sequence of positively and negatively sloped segments. The descriptor sequence is defined by the sign s of the first segment and the total number of segments L and is used to correlate segments across harmonics. For other species (e.g., blue jay and American crow), a predictable sequence of slopes does not exist in the data analyzed.

C. The feature vectors

The feature vectors for each vocalization class were determined by careful analysis of the track sets derived for training. For different species, different feature sets were required to obtain reasonably good classification results.

A single track gives rise to two classes of features: track-derived features v_A and segment-derived features v_B; these are used to form feature vectors $v=(v_A,v_B)$ for each species. A SNR η is also derived from the track and used for thresholding purposes during recognition. In the following, let $\mathcal{T}$ be a track set and $T \in \mathcal{T}$ be a track.

1. Herring gull and Canada goose

A. Let $v_A=(\check{f},\hat{f},\bar{f},\tau,A)$, where $\check{f}$=minimum track frequency; $\hat{f}$=maximum track frequency; $\bar{f}$=mean track frequency; τ=track time length; and A=track maximum amplitude normalized by track set maximum amplitude.

The track SNR η is computed under the assumption that harmonic signal energy is sparse in the spectrogram representation, allowing for a good estimate of noise level using a median filter. For each frame time, a coarse grid is formed at integer multiples of frequency interval size W_f. The noise level in the grid interval from $l \times W_f$ to $(l+1) \times W_f$, where l is a non-negative integer, is estimated by the median value of all spectrogram points in that interval. The ratio η is computed by dividing the track maximum amplitude by the noise level from the grid intervals containing the frequency $(\hat{f}+\check{f})/2$ averaged over all the track time frames. Interval size W_f=500 Hz was used providing roughly twelve 43 Hz bins, adequate for the median filter assuming harmonic tracks occupy three or fewer bins.

B. Herring gull and Canada goose each possess a descriptor sequence S. For herring gull $S=(-,1)$, i.e., the sequence is one negatively sloped segment, and $v_B=(\tau_1,\delta_1)$; for Canada goose $S=(+,2)$, i.e., a positively sloped segment followed by a negatively sloped segment, and $v_B=(\tau_1,\delta_1,\tau_2,\delta_2)$. Here, τ_i and δ_i are the segment duration and slope corresponding to segment i in the descriptor sequence. The tracks in $\mathcal{T}$ can contain more segments than the length of the descriptor sequence; only the duration and slope of the longest matching segment(s) to the descriptor sequence are retained in v_B.

2. Blue jay

A. The track-derived feature vector $v_A=(\check{f},\hat{f},\bar{f},\tau,A)$ is the same as for herring gull and Canada goose.

B. For blue jay, segment slope sign is not consistent, and a descriptor sequence was not identified. Let $v_B=(\check{\delta},\hat{\delta},\bar{\delta})$, where $\check{\delta}$=minimum segment slope; $\hat{\delta}$=maximum segment slope; and $\bar{\delta}$=median segment slope.

3. American crow

Based on the data analysis, crow tracks can be qualitatively described as oscillations superimposed on a concave curve, with many local extrema. This makes segmentation difficult and the American crow model uses only track-derived features. The main issue is capturing the concavity in a robust way. Critical to this is finding a local frequency maximum, which is not at a track endpoint. Of the three greatest, such maxima choose the one closest to the track time midpoint. If it exists, call this point $(t_C,\hat{f}_C)$. Let $(t_L,\check{f}_L)$ be the absolute minimum frequency track point furthest to the left of $t=t_C$, and $(t_R,\check{f}_R)$ be the minimum such point furthest to the right. Let $\rho=(t_C-t_L)/(t_R-t_C)$ (a skewness measure), $\Delta f_L=\hat{f}_C-\check{f}_L$, and $\Delta f_R=\hat{f}_C-\check{f}_R$. The American crow feature vector is then $v_A=(\check{f}_L,\hat{f}_C,\check{f}_R,\rho,\Delta f_L,\Delta f_R,\tau)$, where τ is duration. The SNR η is computed as above.

D. Statistical model

The tracks used to derive features capture the harmonic structure for the vocalization type. In general, there will be $K_h \leqslant K$ tracks for harmonic h in the training data for vocalization V, as not all instances will contain all the harmonics due to vocalization variability. Each such track results in a feature vector v_i, according to species type, and a SNR η_i. Feature statistics are estimated by computing the mean $\bar{v}_h$ and the covariance matrix S_h of the features over the K_h tracks for each harmonic h, resulting in a total of H mean-vector and covariance-matrix pairs (where H is the number of harmonics in the model). Denote the Mahalanobis [15] distance between feature vector v and feature vector mean $\bar{v}$ by $D_h(v)$. For each harmonic h, compute the maximum such distance and the minimum SNR in the training set

$$\hat{D}_h = \max_{i=1}^{K_h} D_h(v_i), \tag{2}$$

$$\eta_h = \min_{i=1}^{K_h} \eta_i. \tag{3}$$

To summarize, each vocalization model consists of some number of harmonics, and a descriptor sequence (which might be null). Each harmonic is described by a Mahalanobis distance function $D_h(\cdot)$, a distance threshold $\hat{D}_h$, and a SNR threshold $\check{\eta}_h$. In matching test vocalization harmonics to a particular model, the test SNR η must satisfy $\eta > \check{\eta}_h$, and the features must fall within the distance $\hat{D}_h$ for h. For each species, some minimum number of such harmonic matches

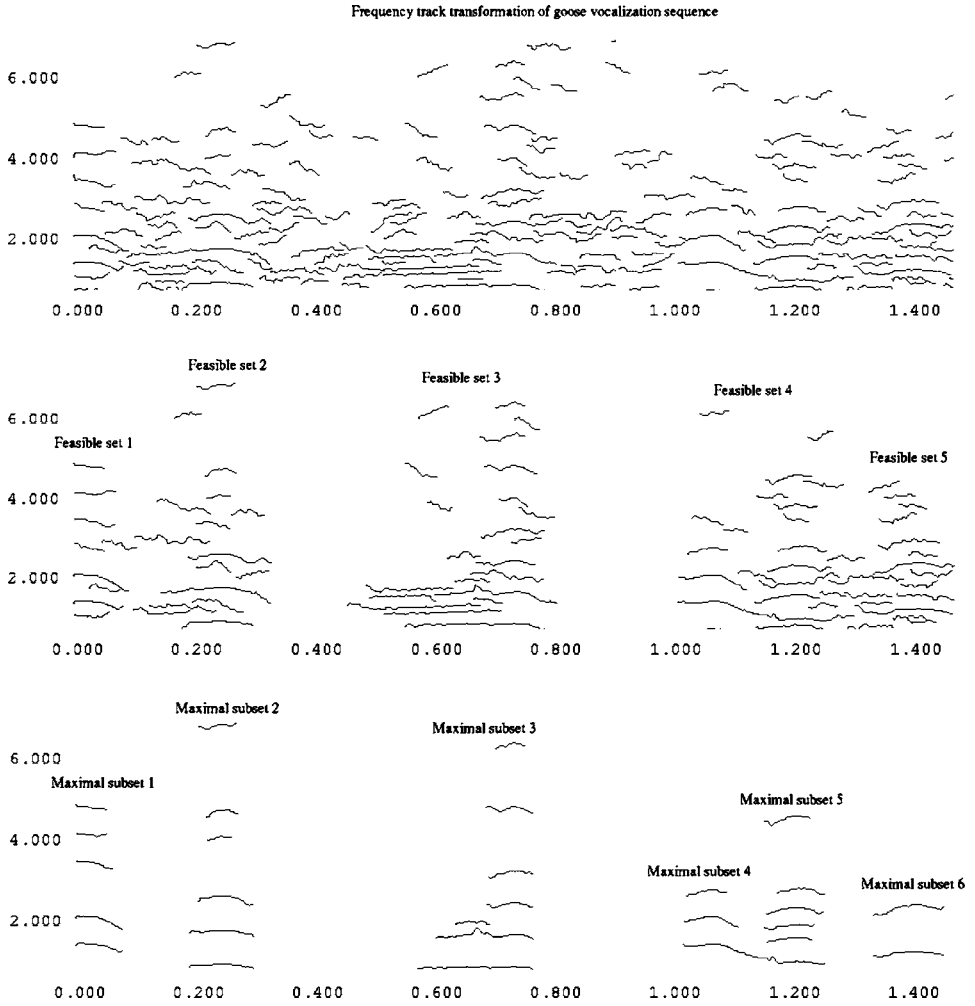


FIG. 2. Image of all tracks in a track set of consecutive goose vocalizations, followed by image of the five feasible sets capturing the vocalizations, followed by the image of the six maximal subsets calculated from the feasible sets. Note that FMS splits feasible set 4 into the two maximal subsets maximal subset 4 and maximal subset 5.

must occur for classification to occur. These numbers are any 3 out of 6 for herring gull, any 4 out of 10 for blue jay, harmonic 1 and at least one other out of 8 for Canada goose (since harmonic 1 occurred in every example in the training set), and 1 out of 1 for American crow. The numbers of required harmonics for the blue jay and herring gull were selected to eliminate false positives. When only three harmonics for blue jay or two harmonics for herring gull are required, false positives occur in the test data of other bird vocalizations.

IV. RECOGNITION

Recognition is accomplished in three steps: transformation of the input recording into a track file using the FTE algorithm, extraction of track sets from it that represent possible harmonic sounds, and comparison of each extracted set with the previously determined bird vocalization models.

A. Track set extraction

Track set extraction is done using two operations: an overlap criterion algorithm called find feasible sets (FFS) followed by a harmonic criterion algorithm called find maximal sets (FMS); see Fig. 2.

1. Find feasible sets

Denote by $\mathcal{T}=\{T_1,\ldots,T_N\}$ the track set resulting from an input recording of arbitrary length. For each track $T_i \in \mathcal{T}$ form the corresponding *accreted set* $\mathcal{A}_i \subset \mathcal{T}$ where $T_j \in \mathcal{A}_i$ if T_j overlaps T_i sufficiently in time, i.e., time overlap is greater than δ_t, where δ_t is the minimum allowed track length as set in track pruning. The accreted set $\mathcal{A}_i$ contains all tracks potentially harmonically related to track T_i plus others due to noise, other vocalizations, and interfering signals. Each track is assumed to belong to at most one vocalization. A set function E that maps pairs of track sets into new pairs is used to obtain track sets $\mathcal{A}_i' \subset \mathcal{A}_i$ such that $\cup_i \mathcal{A}_i' = \mathcal{T}$ and $\mathcal{A}_i' \cap \mathcal{A}_j' = \emptyset$, for all i, j. The function, called *pair erosion*, is defined by

$$E(\mathcal{A}_i,\mathcal{A}_j) = \begin{cases} (\mathcal{A}_i - \mathcal{A}_i \cap \mathcal{A}_j, \mathcal{A}_j) & \text{for } I(\mathcal{A}_i,\mathcal{A}_j) < 0 \\ (\mathcal{A}_i, \mathcal{A}_j - \mathcal{A}_i \cap \mathcal{A}_j) & \text{for } I(\mathcal{A}_i,\mathcal{A}_j) \geq 0. \end{cases} \tag{4}$$

Here, $I(\mathcal{A}_i,\mathcal{A}_j)$ is a measure of the *importance* of (accreted) track set $\mathcal{A}_i$ relative to track set $\mathcal{A}_j$, defined by

$$I(\mathcal{A}_i,\mathcal{A}_j)=\left(\frac{N_i}{N_j}+\frac{\tilde{A}_i}{\tilde{A}_j}\right)-\left(\frac{N_j}{N_i}+\frac{\tilde{A}_j}{\tilde{A}_i}\right), \tag{5}$$

where N_i is the number of tracks in $\mathcal{A}_i$ and $\tilde{A}_i$ is the average of the N_i track maximum amplitudes. If $I(\mathcal{A}_i,\mathcal{A}_j)>0$, then we say $\mathcal{A}_i$ is more important than $\mathcal{A}_j$. The scoring function I can be used to establish an order relation on a set of track sets.

Pair erosion is the removal of duplicate tracks from the less important track set. It is applied to each pair of accreted sets using the mapping $(\mathcal{A}_i,\mathcal{A}_j)\rightarrow E(\mathcal{A}_i,\mathcal{A}_j)$ until the resulting sets are pairwise disjoint. The remaining sets are called *feasible sets*. Some of the feasible sets will be empty; these are discarded. Although E is not quite symmetric, it is unlikely for two sets to have equal importance due to the floating point values used in its computation. More significantly, it is possible for the feasible sets to depend on the order of the application of erosion. The time order imposed by the start of the earliest track in an accreted set was used here; this is consistent with the real-time flavor of the algorithm. Finally, note that equal weighting is given to the terms N_i/N_j compared to $\tilde{A}_i/\tilde{A}_j$; deeper investigation might reveal that this is not optimal.

2. Find maximal sets

The algorithm FMS is then applied iteratively to each feasible set. It uses the harmonic correlation between two tracks to determine whether or not the tracks are related. Two tracks are harmonically correlated when the frequency ratios on the overlapping portions are nearly constant. A measure of this is the function *harmonic relate* defined by

$$H(T_j,T_k)=\frac{1}{M\bar{r}}\sum_{i=1}^{M}|\bar{r}-r_i|. \tag{6}$$

Here, M is the number of pairs of track points (t_i,f_i,A_i) and (t_i,g_i,B_i) in the overlapping portions of the two tracks T_j and T_k. To make H symmetric, T_j and T_k are ordered so that $\bar{f}_i<g_i$. Then, $r_i=g_i/f_i$ and $\bar{r}=\Sigma_{i=1}^{M}r_i/M$.

Let $\mathcal{A}_i$, $i=0,1,\ldots$ be a track set with $\mathcal{A}_0$ a nonempty feasible set. For each track T_j in $\mathcal{A}_i$, compute the *harmonic accreted set* $\mathcal{H}_j$ containing all tracks T_k such that $H(T_j,T_k)<\epsilon$, where $\epsilon>0$ is a user-specified threshold value; for this work, $\epsilon=0.01$. After ordering by importance, the most important harmonic accreted set is relabeled $\bar{\mathcal{H}}_i$, and we derive an eroded feasible set $\mathcal{A}_{i+1}=\mathcal{A}_i-\bar{\mathcal{H}}_i$. The process is iterated until the resulting eroded set is empty, yielding the *maximal sets* (short for *maximal importance harmonic accreted sets*) $\bar{\mathcal{H}}_k,k=0,\ldots,n-1$. For an isolated vocalization, FMS generally produces one large maximal set $\bar{\mathcal{H}}_0$ and several small ones consisting of noise products; for coincident vocalizations, it can separate the corresponding track sets.

B. Comparing extracted sets with the model

To summarize, each input recording for recognition is converted into a track set using FTE, then partitioned in time into feasible sets, which are then partitioned in frequency into harmonically related maximal track sets. Each maximal set $\bar{\mathcal{H}}_k$ is compared with the model of each species vocalization by comparing the tracks in it to the harmonic models in the vocalization model. If the track feature vector falls within the Mahalanobis distance of the submodel for that harmonic track, it is called a *harmonic match*. If the set contains sufficiently many harmonic matches (see above), the recognizer declares the set $\bar{\mathcal{H}}_k$ an occurrence of the vocalization.

TABLE III. Test data.

Species	No. of vocalizations
American crow	52
Blue jay	45
Canada goose	45
Herring gull	132
Northern cardinal	4
Fish crow	38
House sparrow	47
Mallard	45
Nuthatch	15
European starling	30
Barn swallow	>200
Carolina wren	>150
Eastern meadowlark	>90
Indigo bunting	>90
Painted bunting	>250
Pine warbler	>40
Red-winged blackbird	>100
Savannah sparrow	>100
Scissor-tailed flycatcher	>130
Tufted titmouse	>300

C. Examples of recognition of target species

The method just described was applied to blue jays, herring gulls, and Canada geese. Track set extraction was not needed for American crows since the vocalization model contains only one track. Both rejection and detection performance were measured. Two test data sets were used: one collected by the authors containing the target and other species and the other collected by Benson containing nontarget species. Benson's data set was used purely to test the ability of the recognition algorithm to reject sounds other than the target vocalization type. Ten species (including the four target species) were selected from our data set. The six other species in our data set were northern cardinal (*Cardinalis cardinalis*), fish crow (*Corvus ossifragus*), house sparrow (*Passer domesticus*), mallard (*Anas platyrhynchos*), white-breasted nuthatch (*Sitta carolinensis*), and European starling (*Sturnus vulgaris*); these recordings were also used to test rejection performance of the algorithm. Ten species were also selected from Benson's data set: barn swallow (*Hirundo rustica*), Carolina wren (*Thryothorus ludovicianus*), eastern meadowlark (*Sturnella magna*), indigo bunting (*Passerina cyanea*), painted bunting (*Passerina ciris*), pine warbler (*Dendroica pinus*), red-winged blackbird (*Agelaius phoeniceus*), savannah sparrow (*Passerculus sandwichensis*), scissor-tailed flycatcher (*Tyrannus forficatus*), and tufted titmouse (*Baeolophus bicolor*). Table III lists the exact number

TABLE IV. Recognition results.

Species	No. of vocalizations	No. of correctly identified
Blue jay	45	37
Herring gull	16	15
Canada goose	45	37
American crow	52	36

of vocalizations in the recordings of the target species and the approximate number of vocalizations in the recordings of the other 16 species.

1. Blue jay

The training set consisted of 64 hand-pruned track files corresponding to 64 blue jay vocalizations from 20 short recordings, recorded in summer, all from the same individual. The test set consisted of 45 vocalizations, 19 of from the same individual as the training set and 26 from several other blue jays recorded at a nearby location in winter. The principal difficulties observed for blue jay calls were that they became chaotic, resulting in broken and irregular tracks at vocalization end, and the overall slope characteristics of the call tracks varied from call to call, resulting in a null descriptor sequence. 37 out of 45 blue jay vocalizations were correctly identified (see Table IV); no false positives resulted from other species.

2. Herring gull

Herring gulls vocalize with a variety of syllables, each requiring its own feature model. The syllable types are distinguished from each other by the number and length of the segments, the descriptor sequence, and the vocalization fundamental frequency. In the recording database, there is one syllable that occurs 31 times, and several others that occur 7 or fewer times each. A model for the most common syllable was made using 15 of the 31 processed track files for a training set. Since the syllable is always downsloping in frequency, its descriptor sequence is $(-, 1)$. Using the values of $\hat{D}_h$ determined during training, applied to the recordings containing the 16 remaining syllables, 11 were correctly recognized. This low rate is likely due to the limited amount of training data, which produced a Mahalanobis distance threshold less than 14 on each harmonic, as compared to 40 for blue jay and 38 for Canada goose. Using a value of 40 for each harmonic resulted in 15 out 16 correctly identified with no false positives, as shown in Table IV. Increasing this value further, there were no false positives up to a distance value of 75 and all 16 vocalizations were correctly recognized. At a value of 100, there were false positives for eastern meadowlark and other gull vocalization syllables.

3. Canada goose

Canada geese have one primary vocalization type and several secondary vocalizations. This work focused on recognizing only the primary vocalization type. The frequency curve of the primary vocalization type is concave down with an upsloping part followed by a downsloping part [so the descriptor sequence is $(+,2)$]. Less commonly, Canada geese make a similar vocalization with no upsloping initial part, as well as hissing and grunting sounds. The training data used to make the model for the primary goose vocalization type was collected from a large flock of geese at Sunken Meadow State Park and the data (a 1 min long recording of a flock) used to test recognition performance of this vocalization type was collected at the park later on the same day. Many different birds were vocalizing in both the training and test set, and a fairly wide range (500–850 Hz) of vocalization fundamental frequencies are present. The test set contained 45 primary goose vocalizations, of which 37 were correctly recognized by the algorithm. The recognizer was again applied to the data set of other species, with a false positive occurring in one of the American crow recordings (as shown in Table V).

TABLE V. Rejection results.

Species	No. of false positives	Type
American crow	3	Canada goose
Canada goose	1	American crow

4. American crow

For American crows, recognition is carried out track by track, and track set extraction is not done. Moreover, the crow model is considerably different from the other three species. In particular, the Mahalanobis distance criterion produced a large number of false positives, and an alternative distance function was used. Using a training set containing 40 crow vocalizations and a test set of 52 crow vocalizations, the results in Tables IV and V were obtained. These results are included here for completeness. The measure used was binary: The maximum and minimum values for each feature in the crow training set were used to determine a feature range. A track matched the harmonic model if each feature fell within the range.

V. CONCLUSION

This paper demonstrates an algorithm for harmonic frequency track extraction and its application to classification of birds by vocalization analysis. First, a frequency track representation of the sound segment is calculated, resulting in a track file. Second, the track file is separated into track sets (feasible sets) calculated using an overlap criterion. Third, each feasible set is divided into disjoint harmonically related subsets (maximal sets). Finally, each maximal set is compared with the vocalization models.

The performance of the recognizer was fair (crow) to very good (blue jay) on target species, and excellent on false positives. Improved recognition performance can likely be obtained by more refined feature sets and more extensive data sets. The only false positives that occurred were when crow and goose were identified as one another. If more spe-

cies were considered, it is possible that more features would be needed in order to achieve this level of performance in rejecting false positives.

ACKNOWLEDGMENTS

This work was funded by the Air Force Office of Scientific Research, the Research Foundation of Stony Brook University, and Advanced Acoustic Concepts, Inc. The authors would like to thank Willard Larkin of AFOSR for suggesting the application of acoustic processing to bird-strike avoidance; to James Glimm, Bruce Stewart, and Robert Benson for helpful comments and discussions; and to the reviewers for their constructive criticisms.

[1]E. C. Cleary, R. A. Dolbeer, and S. E. Wright, "Wildlife strikes to civil aircraft in the United States: 1990–2005," Technical Report 12, Federal Aviation Administration, Washington, DC, 2006.

[2]R. P. DeFusco, "A bird avoidance model for the US Air Force," USAFA Discovery Magazine **98**, 1–2 (1998).

[3]R. Shade, "Auditory cortex analysis phase 2 proposal," Technical Report, Advanced Acoustic Concepts, Hauppauge, NY, 2004.

[4]J. A. Kogan and D. Margoliash, "Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study," J. Acoust. Soc. Am. **103**, 2185–2196 (1998).

[5]A. Härmä and P. Somervuo, "Classification of the harmonic structure in bird vocalization," in IEEE ICASSP (2004), pp. 17–21.

[6]S. Fagerlund, "Automatic recognition of bird species by their sounds," Ph.D. thesis, Helsinki University of Technology, Helsinki 2004.

[7]Z. Chen and R. C. Maher, "Semi-automatic classification of bird vocalizations using spectral peak tracks," J. Acoust. Soc. Am. **120**, 2974–2984 (2006).

[8]S. E. Anderson, A. S. Dave, and D. Margoliash, "Template-based automatic recognition of birdsong syllables from continuous recordings," J. Acoust. Soc. Am. **100**, 1209–1219 (1996).

[9]J. R. Buck and P. L. Tyack, "Quantitative measure of similarity for Tursiops truncatus signature whistles," J. Acoust. Soc. Am. **94**, 2497–2506 (1993).

[10]D. K. Mellinger and C. W. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," J. Acoust. Soc. Am. **107**, 3518–3529 (2000).

[11]L. R. Rabiner and B. Juang, *Fundamentals of Speech Recognition* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

[12]K. M. Fristrup and W. A. Watkins, "Marine animal sound classification," Technical Report 13, Woods Hole Oceanographic Institution, Woods Hole, MA, 1994.

[13]R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Process. **34**, 744–754 (1986).

[14]X. Serra and J. Smith, "Spectral modeling synthesis," Comput. Music J. **14**, 12–24 (1990).

[15]K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis* (Academic, New York, 2000).

Syllable acoustics, temporal patterns, and call composition vary with behavioral context in Mexican free-tailed bats

Kirsten M. Bohn[a)]
Section of Neurobiology, University of Texas at Austin, Austin, Texas 78712

Barbara Schmidt-French
Bat Conservation International, 500 North Capital of Texas Highway, Building 1, Suite 200, Austin, Texas 78746

Sean T. Ma
Institute for Neuroscience, University of Texas at Austin, Austin, Texas 78712

George D. Pollak
Section of Neurobiology, University of Texas at Austin, Austin, Texas 78712

(Received 21 November 2007; revised 22 May 2008; accepted 7 June 2008)

Recent research has shown that some bat species have rich vocal repertoires with diverse syllable acoustics. Few studies, however, have compared vocalizations across different behavioral contexts or examined the temporal emission patterns of vocalizations. In this paper, a comprehensive examination of the vocal repertoire of Mexican free-tailed bats, *T. brasiliensis*, is presented. Syllable acoustics and temporal emission patterns for 16 types of vocalizations including courtship song revealed three main findings. First, although in some cases syllables are unique to specific calls, other syllables are shared among different calls. Second, entire calls associated with one behavior can be embedded into more complex vocalizations used in entirely different behavioral contexts. Third, when different calls are composed of similar syllables, distinctive temporal emission patterns may facilitate call recognition. These results indicate that syllable acoustics alone do not likely provide enough information for call recognition; rather, the acoustic context and temporal emission patterns of vocalizations may affect meaning. © *2008 Acoustical Society of America.*
[DOI: 10.1121/1.2953314]

PACS number(s): 43.80.Ka [JAS] Pages: 1838–1848

I. INTRODUCTION

Extensive research on vocal communication has been conducted on anurans, birds, cetaceans, and nonhuman primates (Todt *et al.*, 1988; Tyack and Clark, 2000; Ryan, 2001; Marler, 2004). Less attention has been given to bats although several studies report various types of communication calls. These include infant isolation calls (e.g., Gould *et al.*, 1973; Gelfand and McCracken, 1986), maternal directives (Esser and Schmidt, 1989; Balcombe and McCracken, 1992), mating songs (Barlow and Jones, 1997; Davidson and Wilkinson, 2002; Behr and von Helversen, 2004), distress calls (Russ *et al.*, 1998; Russ *et al.*, 2004), and group foraging calls (Wilkinson and Boughman, 1998). A handful of studies have also examined multiple vocalizations within species. These studies have revealed rich repertoires (Barclay *et al.*, 1979; Porter, 1979; Kanwal *et al.*, 1994; Pfalzer and Kusch, 2003; Ma *et al.*, 2006; Melendez *et al.*, 2006), including rules for syllable combinations (Kanwal *et al.*, 1994). Vocalizations can be quite intricate (Behr and von Helversen, 2004), and some species are capable of vocal learning (Esser, 1994; Boughman, 1998). Thus, bats are an intriguing taxon for vocal communication research but studies that incorporate behavioral and vocal diversities within a species are rare.

Examining multiple vocalizations and corresponding behaviors within a species not only provides an overview of vocal diversity but also permits comparisons of acoustic features across different behavioral contexts. If vocalizations are correlated with different behavioral contexts, they likely have different meanings that should be reflected in their acoustics. Calls are usually identified and described by the acoustic features of individual elements (e.g., Cleveland and Snowdon, 1982; Seddon *et al.*, 2002; Range and Fischer, 2004), which we will refer to as syllables. Syllable acoustics can also vary with and provide information on identity at the individual (Cheney and Seyfarth, 1980; Hauser, 1991; Bohn *et al.*, 2007), group (Boughman, 1997), or family (Gouzoules and Gouzoules, 1990) level.

Although syllable acoustics may be an important means for conveying information, other features of vocalizations may be equally significant. There are at least four other ways in which vocalizations have varied with behavioral context or elicited differential responses from receivers. (1) The temporal distribution of syllable emissions. Most salient is whether syllables are emitted in distinct clusters (i.e., calls are multisyllabic) or emitted singly (i.e., calls are monosyllabic; Cleveland and Snowdon, 1982). (2) The rate at which syllables are emitted over time (i.e., call rate; Hauser, 1998;

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

Ghazanfar *et al.*, 2002; Searby *et al.*, 2004). (3) The number of syllables per call (Ghazanfar *et al.*, 2002). (4) Acoustic context, that is, whether and how syllables are combined (Cleveland and Snowdon, 1982; Robinson, 1984; Crockford and Boesch, 2005). The first two features, temporal distribution and call rate, relate to temporal emission patterns whereas the second two features, the number of elements per call and acoustic context, relate to call composition. Both temporal emission patterns and call composition have received much less attention than syllable acoustics, particularly in bats, even though they may be perceptually salient and an integral component of vocal communication.

Here, we examine the vocal repertoire of Mexican free-tailed bats and test whether syllable acoustics, temporal emission patterns, and/or call composition vary among vocalizations emitted in different behavioral contexts. We describe 16 vocalizations associated with particular behaviors, including complex territorial-courtship song produced by males during the mating season. For each vocalization, we examine the distribution of syllable emissions over time and use a quantitative analysis to determine whether vocalizations consist of single syllables or whether multiple syllables are emitted as units. Finally, we determine whether different calls are distinguishable based on syllable acoustics, temporal emission patterns, or call composition features.

II. METHODS

A. Study site and animals

This study was conducted on a captive colony of approximately 60 *T. brasiliensis*. The colony has been maintained by Schmidt-French in Austin, TX, for a period of ten years, and the identity, sex, and history for each individual have been documented. Schmidt-French could identify individuals by sight or after examination. Bats were housed in a wooden structure measuring 4.9 m (length) $\times$ 3.7 m (width) $\times$ 3.7 m (height). Two windows allowed filtered sunlight to enter. Humidity was maintained at 60% or above and temperatures varied in the building from approximately 22 to 26 °C. Cloth-covered heating pads placed in cages during evening hours provided bats with the option of accessing temperatures reaching 29 °C. The bats roosted in fabric pouches positioned along the walls and ceilings of open wooden cages and had access to the entire building. Bats had continual access to water and beetle larvae (*Tenebrio molitor*), and were also offered a blended mixture of larvae, baby food, and vitamin supplements in the evening (Lollar and Schmidt-French, 2002).

B. Acoustic recordings

We observed behaviors and recorded vocalizations during the late evening (2200–0100 h) when the bats were most active. The bats were accustomed to the building being well lit during this time allowing direct observation and recording of behaviors onto a digital camcorder (Sony DCR-TRV-460). Behaviors that occurred inside roosting pouches were recorded using a Night Owl Cam (Model NI 201), a complementary metal oxide semiconductor camera with infrared and microphone (Birdhouse Spy). In order to compare social vocalizations with echolocation calls, we recorded echolocation calls from each of the three bats while they flew in the colony room. Vocalizations were recorded using a $\frac{1}{4}$ in. microphone (Brüel and Kjær type 4939) and a custom-made amplifier. In 2003, signals were recorded into a custom-made digital time expander. The time expander recorded a maximum of 10 s at 16 bits that was played onto a computer at a sample rate of 44.1 kHz. In 2004 and 2005, calls were recorded directly onto a computer at a sample rate of 300 kHz using a high-speed data acquisition card (National Instruments, NI PCI 6251 M Series, Austin, TX) and AVISOFT software (version 2.97, Avisoft Bioacoustics, Berlin, Germany). Both systems allowed recordings up to 150 kHz, well above the frequency content of vocalizations. For analyses, we only examined recordings where the identity of the caller was known and the corresponding behavior was observed. After examining vocalizations and recordings, we further restricted our analyses to those vocalizations that were unambiguously correlated with specific behaviors and observed in at least three individuals.

C. Terminology

We used the following terms to describe vocalizations:

Syllable. The smallest acoustic unit of a vocalization in this study and equivalent to one continuous emission surrounded by silence (Kroodsma, 1977; Kanwal *et al.*, 1994).

Call. The simplest emission of a vocalization. Calls can be composed of single syllables or groups of syllables. For all vocalizations except for song, we use quantitative methods (see Sec. II D) to determine what constitutes a call. If syllables are emitted singly, then each syllable is a *monosyllabic* call. Alternatively, if multiple syllables are always emitted together then the group of syllables is a *multisyllabic* call.

Bout. A group of calls.

Song. Vocalizations emitted by males during the mating season that have multiple types of syllables and phrases (Catchpole and Slater, 1995; Marler and Slabbekoorn, 2004).

Phrase. A combination of one or more types of syllables that may be repeated in a song (Kroodsma, 1977; Marler and Slabbekoorn, 2004). *Simple phrases* are composed of one type of syllable. *Complex phrases* are composed of different types of syllables (similar to a "note complex" phrase in birds, Kroodsma, 1977; Marler and Slabbekoorn, 2004).

D. Temporal analysis

We used temporal analyses to classify calls as mono- or multisyllabic and to demarcate intracall intervals from intercall intervals. For all recordings, we normalized amplitudes and determined the beginning and end times of syllables on oscillograms. We tested whether syllables were distributed randomly across time by dividing recordings into equal time intervals and then comparing the distribution of the number of syllables per time interval with a Poisson distribution. Time interval lengths varied by call but were selected so that the mean number of syllables per interval was consistently between 1 and 2 (Zar, 1984). If the temporal distribution of syllables was significantly different than random, we classi-

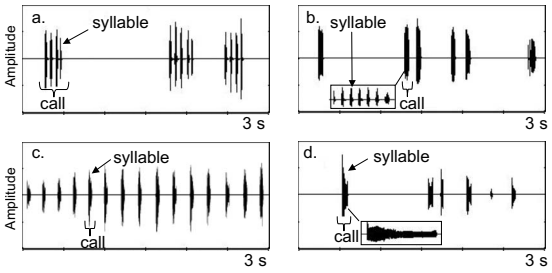


FIG. 1. Temporal emission patterns of *Tadarida* vocalizations. (a) and (b) are multisyllabic calls with clumped syllable distributions and syllables are never emitted singly. (c) Monosyllabic calls with a uniform syllable distribution and (d) monosyllabic calls with a random distribution. (a) Maternal directives, (b) irritation calls, (c) mounting calls, and (d) marking calls.

fied the distribution as clumped if the variance was greater than the mean, or uniform if the variance was less than the mean (Zar, 1984). We considered calls multisyllabic if syllables were emitted in a clumped manner where each "clump" represented a call [Figs. 1(a) and 1(b)], and monosyllabic if syllables were emitted randomly or uniformly [Figs. 1(c) and 1(d)]. In order to avoid biasing results by including long intervals of silence, we did not use intervals of 4000 ms or more (32 of 5726 intervals).

Next, for calls with clumped distributions, we used bout analyses (Sibly *et al.*, 1990) to determine thresholds between intercall intervals and intracall intervals (intervals between syllables within calls). For each call type, we calculated the log frequency of intervals per unit time across the range of interval lengths. We used nonlinear regression (PROC NLIN, SAS Institute, Cary, NC) to fit a two-process model to the data, where the fast process represented syllable intervals and the slow process represented call intervals (Sibly *et al.*, 1990). Once intercall thresholds were established, we calculated the number of syllables per call for each call type.

E. Syllable measurements

We measured acoustic features for syllables from each call and for four types of syllables found in song. We randomly selected five syllables from three randomly selected bats for each syllable type. We band-pass filtered syllables between 1 and 100 kHz and normalized amplitudes by dividing each signal by its peak amplitude. Syllables were measured using SASLAB PRO (version 4.39, Avisoft Bioacoustics). We measured syllable duration (dur) from oscillograms. We measured five frequency characteristics of the fundamental frequency, which was usually dominant: peak frequency (f_{peak}), beginning frequency (f_{beg}), end frequency (f_{end}), maximum frequency (f_{max}), and minimum frequency (f_{min}). Peak frequency (f_{peak}) was measured from power spectrums and the other frequency variables were measured from spectral contours. Click syllables and protest syllables were measured differently. Clicks were of such short duration (<1 ms) that we could not construct meaningful spectral contours and so we did not measure f_{beg} and f_{end}, and we measured f_{min} and f_{max} on power spectrums at -20 dB below peak frequency. Protest syllables had most energy at harmonics above the fundamental and the fundamental often did not clearly show the shape of the syllables. For these calls, we measured f_{peak} at the fundamental but measured the remaining four frequency variables from the dominant harmonic. For descriptive purposes alone, we also calculated call bandwidths as $\log_2(f_{max}/f_{min})$.

F. Syllable acoustic analysis

For syllable analyses, our goals were to (1) describe syllable acoustic structure and variation and to (2) determine whether syllable acoustics differed when syllables were incorporated into different calls or song. First, we visually examined syllables and observed that they fell into three acoustic groups: clicks, downward frequency-modulated (FM) sweeps (down FMs), and long syllables. We used discriminant function analysis to confirm these groupings using f_{peak} and dur, as these were the only variables measured consistently across all syllables. Next, for each group, we screened variables for high correlations. We removed highly correlated measurements. We excluded maximum (f_{max}) and minimum (f_{min}) frequencies in analyses for down FMs because they were highly correlated with beginning (f_{beg}) and end (f_{end}) frequencies, respectively ($r=0.99$, $p<0.0001$, $N=115$ for both pairs of variables). Next, within each of the three groups, we used a multiple analysis of variance (MANOVA) with call as a factor and bat nested within call as a random factor to determine if syllables differed among calls or bats (PROC GLM, SAS). For acoustic groups that showed significant differences between calls, we performed discriminant function analyses to determine which syllables varied among different calls (PROC DISCRIM, SAS). For all discriminant function analyses and MANOVAs, we tested for normality using Shapiro–Wilk tests and, if necessary, used $\log_{10}$ and square-root transformations to meet normality requirements. Misclassification rates were calculated using leave-one-out cross-validation (PROC DISCRIM, SAS). We present figures of canonical functions with eigenvalues greater than 1. All temporal variables are presented in ms and all frequency variables in kHz.

G. Temporal emissions and call composition

In the next analysis, we determined whether calls that were composed of similar syllables differed in their temporal emission patterns or call composition features. First, we compared the distribution of syllable emissions. Multisyllabic calls (calls with clumped distributions) were considered different and discriminable from monosyllabic calls (calls with random or uniform distributions). Second, we considered the acoustic context of vocalizations and considered cases where syllables were embedded in complex phrases in song as discriminable from other calls. Third, we compared syllable emission rates. For monosyllabic calls, we tested whether call rates (calls/s) differed across calls using analysis of variance (ANOVA) and Tukey tests (Zar, 1984). Call rates were calculated for each recording but there were not enough recordings for each bat to include it as a nested factor. For multisyllabic calls, we used ANOVAs with bat as a nested random factor and Tukey tests to compare the interval between syllables within calls. Finally, we compared the number of syllables per call for multisyllabic calls using ANO-

VAs with bat as a nested random factor and Tukey tests. All data are presented as means ± standard error of the mean unless stated otherwise.

III. RESULTS

A. Vocalizations and behaviors

We identified 16 vocalizations associated with different behaviors that we divided into six categories: (1) mating, (2) parent-offspring, (3) antagonistic, (4) social, (5) human interaction, and (6) flight. Below, we briefly describe the behavioral context, acoustic structure, and temporal emission pattern of each vocalization. We also include sample sizes where *bats* and *records* refer to the number of bats we recorded from and the total number of recordings, respectively.

1. Mating

These calls were only emitted by reproductively active adults during the mating season. Most of the year males and females roosted together and males did not establish individual territories nor did they exhibit territorial behavior. In contrast, during the mating season, from January through May, dominant males (approximately 50% of males in the colony) roosted alone in one of the handmade pouches. Each roosting pouch was a dominant male's territory where he actively recruited reproductively active females and fought encroaching males.

Herding calls. These calls were produced during a highly stereotyped behavior, while a male forcefully pushed one or more females with his muzzle or wing into a cluster inside his territory [Fig. 2(a), bats=5, records=14]. Herding calls were multisyllabic (Table I) and consisted of three to ten downward FM sweep syllables (Table II). In approximately 35% of the 116 calls we recorded, syllables were emitted continuously forming sinusoidlike calls [Fig. 2(a)]. In many cases, the down FM sweep calls would transition to sinusoidlike calls within a bout of herding calls.

Marking calls. Dominant males emitted this call while rubbing their faces, gular glands, and penis on the surfaces of their territories [Fig. 2(b), bats=3, records=25]. Males also emitted marking calls over long periods of time (30 min or more) while relatively inactive on their territories, which suggests that marking calls could function in territory maintenance. Marking calls were emitted randomly and considered monosyllabic [Table I and Figs. 1(d) and 2(b)]. Syllables were long (durations up to 80 ms) and low in frequency (4–11 kHz) with highly variable spectrotemporal features.

Mounting calls. Mounting calls were produced during a highly stereotyped behavior. A male would attempt to mount a female, and produce these calls as he forcefully pushed his muzzle repeatedly between her shoulders [Fig. 2(c), bat=8, records=17]. These calls may be associated with dominance, as males never copulated during these encounters and sometimes performed similar behaviors when interacting with juvenile males. Mounting calls were monosyllabic and emitted at regular intervals [Table I and Figs. 1(c) and 2(c)]. Syllables were multiharmonic and relatively constant frequency (bandwidths less than 0.5 octaves).

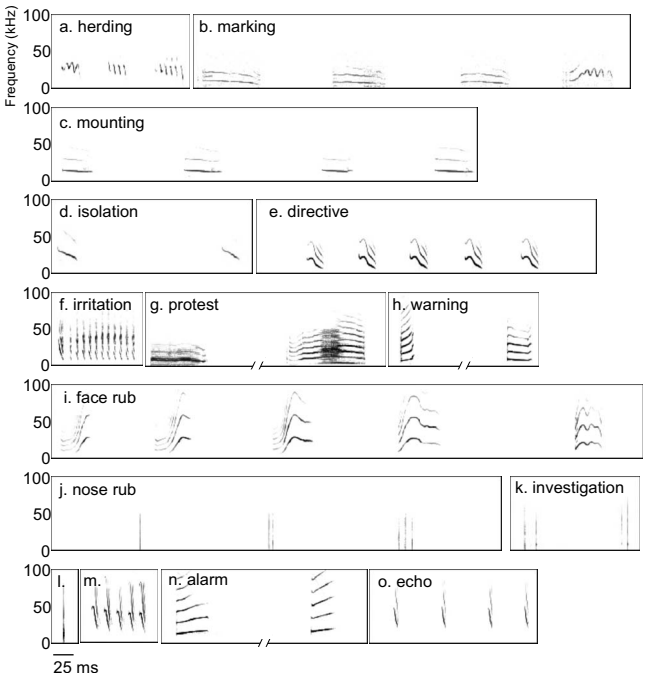


FIG. 2. Spectrograms of calls emitted by *T. brasiliensis*. (a) Three herding calls, (b) four marking calls [also see Fig. 1(d)], (c) four mounting calls [also see Fig. 1(c)], (d) three isolation calls, (e) one directive call [also see Fig. 1(a)], (f) one irritation call [also see Fig. 1(b)], (g) two protest calls (each call is from a different bat), (h) two warning calls (each call is from a different bat), (i) face-rubbing call, (j) nose-rubbing call, (k) investigation clicks, (l) food click, (m) food solicitation call, (n) two alarm calls (each call is from a different bat), and (o) four echolocation calls.

Territorial/courtship songs. Most males that established territories (approximately 75%) produced territorial/courtship song (here forth referred to as "song") when reproductively active females or dominant males approached their territories, or when other males sang. While singing to females, males marked their territories with secretions from their gular glands, anus, and penis, and performed wing-flapping displays (French and Lollar, 1998). Alternatively, if another male intruded on the established territory of a dominant male, the territorial male sang and often chased the intruding male away.

Songs were the most complex vocalizations we recorded and were composed of up to three principal phrases: chirps, trills, and buzzes (Fig. 3, bats=16, records=116). Chirp phrases were complex phrases since they were composed of two types of syllables: "type A" syllables and "type B" syllables [Fig. 3(b)]. Each chirp began with 1–18 repetitions of type A syllables and terminated with a single type B syllable. Type A syllables were brief, 3–7 ms downward FM sweeps. The chirp phrase was repeated 1–36 times in each song (median=7, N=171 songs). Type B syllables were more complex in that they usually had both upward and downward FM components and were longer in duration (14–20 ms, see Sec. III B). In addition to chirps, 60% of songs also contained trills, buzzes, or both phrases [N=171 songs, Figs. 3(c) and 3(d)]. Both trills and buzzes were simple phrases composed of a series of rapidly repeated downward FM sweep syllables.

TABLE I. Results of chi-squared tests on the distribution of syllables over time.

Call[a]	Chi-squared	*df*	*P*	*N* Syllables	Variance/mean	Distribution
Alarm	0.9	3	0.82	103	1.0	Random
Echo	70.4	4	<0.0001	451	0.5	Uniform
Food	610.1	4	<0.0001	403	2.66	Clumped
Herding	186.3	4	<0.0001	448	1.59	Clumped
Isolation	18.7	3	0.0003	553	0.8	Uniform
Irritation	1508.9	5	<0.0001	1203	4.5	Clumped
Marking	2.2	3	0.53	109	0.8	Random
Directive	259.4	4	<0.0001	261	2.9	Clumped
Mounting	44.1	3	<0.0001	120	0.5	Uniform
Protest	4.1	4	0.39	153	0.9	Random
Question	5.9	3	0.12	96	1.1	Random
Warning	77.2	4	<0.0001	249	0.4	Uniform

Food=food solicitation, isolation=infant isolation, and directive=maternal directive. Temporal analyses were not performed on anticipation clicks, nose-rubbing calls, or face-rubbing calls because there were no intercall data. Anticipation clicks were always emitted singly and the other two calls were multisyllabic but only one call was ever recorded at a time.

2. Parent-Offspring

These vocalizations were only produced by lactating females and their young.

Infant isolation calls. Neonates emitted isolation calls immediately after birth and continued to emit these calls throughout development. Pups called when they were isolated or when milk was no longer visible in their stomachs. Infant isolation calls were monosyllabic and although sometimes emitted alone; they were also produced in long trains at relatively constant intervals resulting in a uniform distribution [Table I and Fig. 2(d), bats=4, records=14]. Isolation calls had variable FM patterns that were often simpler and with smaller bandwidths (usually less than one octave) than maternal directives (see below). Previous research has shown that isolation calls are individually distinctive (Gelfand and McCracken, 1986) and that females can discriminate isolation calls from their own pups (Balcombe, 1990).

Maternal directive calls. Maternal directives were emitted throughout pup development when females approached pups or in response to pups' isolation calls. Maternal directive calls were multisyllabic (Table I) with 3–11 syllables per call [Table II and Figs. 1(a) and 2(e), bats=7, records=22]. This call showed a stereotyped number of syllables as demonstrated by a low coefficient of variation (Table II). Directive syllables also had complex FM patterns, often with both upward and downward components that spanned up to two octaves. Previous research has shown that these calls are individually distinctive (Balcombe and McCracken, 1992).

3. Antagonistic

Both males and females, regardless of reproductive status, emitted these calls during agonistic interactions.

Irritation calls. Bats produced these calls when (1) they were jostled by another bat in the roost or (2) when disturbed by a handler. Bats would sometimes push their muzzles into or head-butt another bat while producing this call. In response to irritation calls, bats would often emit protest calls or bare their teeth and emit warning calls. Irritation calls

TABLE II. Results of bout analyses for multisyllabic calls. *R*-squared values for the two-process nonlinear models, intercall threshold from two-process models, median, range, and coefficient of variation (CV=standard deviation/mean) of the number of syllables per call, the interval between syllables within calls (mean ± SE), and sample sizes. Nonlinear models were not fitted to face-rubbing and nose-rubbing calls because we did not have intercall interval data (see text).

Call[a]	R^2	Intercall threshold (ms)	Median syllables	Range syllables	CV[b]	Syllable interval (ms)	*N* calls, *N* intervals
Face rub	···	···	6	4–24	0.64	126.1 ± 6.6	25, 192
Food	0.91[c]	22	3	2–11	0.38	10.6 ± 0.2	153, 291
Herding	0.89[c]	11.5	4	3–10	0.36	3.2 ± 0.1	117, 436
Irritation	0.95[c]	28	5	3–15	0.43	8.7 ± 0.1	237, 1138
Directive	0.83[c]	168	5	3–11	0.29	59.2 ± 1.4	65, 277
Nose rub	···	···	13	5–28	0.49	31.8 ± 2.4	19, 263
Trill	···	···	4	2–10	0.33	3.7 ± 0.1	118, 289
Buzz	···	···	23	5–44	0.45	4.4 ± 0.3	111, 2085

Food=food solicitation, directive=maternal directive, and trill and buzz are from song.
CV=coefficient of variation=(standard deviation/mean) of the number of syllables/call.
$P<0.0001$.

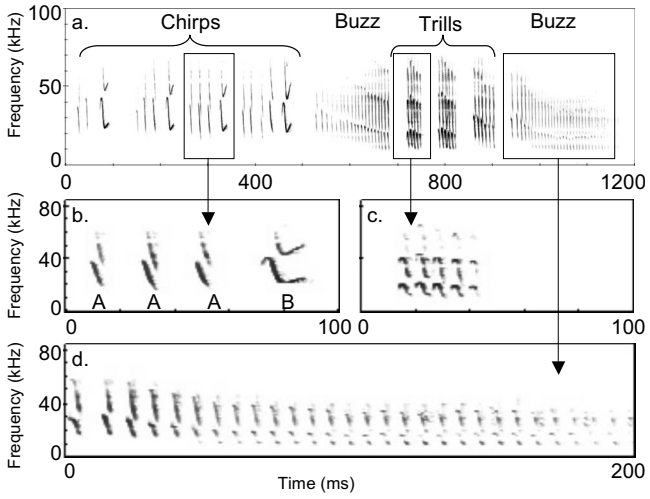


FIG. 3. Territorial/courtship song. (a) One complete song showing the three types of phrases: chirp, buzz, and trill. (b) Expanded section of a chirp, which has two types of syllables: types A and B. (c) Expanded section of a trill. (d) Expanded section of a buzz.

were multisyllabic [Table I and Fig. 1(b), bats=14, records =28], with between 3 and 15 short downward FM sweep syllables per call [Table II and Fig. 2(f)].

Protest calls. We recorded these calls in three situations: when a female struggled to escape a male that was attempting to mount her, when a bat was shoved or swatted by another bat, and when a bat was in discomfort from human handling. When bats emitted protest calls, they behaved submissively; they struggled, retreated, or froze with their bodies flattened against the roost surface. Protest calls were monosyllabic (Table I) and emitted singly or in bouts. Of all vocalizations, protest syllables were the longest (durations up to 105 ms) and had the lowest frequencies [fundamental frequencies between 1 and 2 kHz, Fig. 2(g), bats=14, records=18]. They had variable FM components with bandwidths of the dominant harmonic of up to one octave. Protest syllables were squeal-like in that energy was spread across many harmonics and there were often subharmonics. Dominant harmonics similar to formants occurred at higher frequencies than the fundamental.

Warning calls. We recorded warning calls during three types of situations: when a male entered another male's territory during the mating season, during physical altercations with other bats in the roost, or when in distress from human handling. While emitting these calls, bats would use their heads to shove or used their wings to swat other bats. Bats would respond to warning calls with either aggressive displays (baring teeth and head bobbing) or they would produce protest calls and retreat. Warning calls were monosyllabic and although frequently emitted alone, they were also emitted in long trains at relatively constant intervals resulting in a uniform distribution [Table I and Fig. 2(h), bats=11, records=25]. Syllables were loud and relatively constant frequency with bandwidths less than 0.5 octaves.

4. Social

Both males and females, regardless of reproductive status, emitted these calls while roosting with conspecifics. They were not associated with agonistic interactions.

Face-rubbing calls. Bats emitted face-rubbing calls during a highly stereotyped behavioral sequence. They hopped a few centimeters towards another bat, rubbed their muzzles across the body of the bat, and sometimes tapped the other bat with a folded wing. Face-rubbing calls were multisyllabic since syllables were never emitted alone. We never recorded more than one call at a time with up to 5 min recordings. Consequently, we had no intercall intervals to perform temporal distribution patterns. Face-rubbing calls were composed of 4–24 syllables, but the number of syllables per call was highly variable compared to the other vocalizations (Table II, coefficient of variation=0.64). The structure of face-rubbing syllables was also highly variable [Fig. 2(i), bats=9, records=23]. Spectrotemporal patterns included both upward and downward FM components. Some syllables were relatively constant frequency with bandwidths less than 0.25 octaves, whereas others had bandwidths of up to 3 octaves.

Nose-rubbing calls. These calls were also produced during a specific behavior, while a bat gently rubbed its muzzle against another bat's muzzle. They appeared to possibly function as a greeting because callers usually directed them at bats as they first entered the roost. Nose-rubbing vocalizations were always emitted as multisyllabic calls. We never recorded more than one call at a time (for up to 5 min periods). Thus, like face-rubbing calls, we had no intercall intervals for temporal pattern analyses. Nose-rubbing calls were comprised of between 5 and 28 clicks [Fig. 2(j), bats=9, records=21].

Investigation clicks. We recorded these clicks when a bat approached a novel stimulus at a roost pouch such as an entering bat or human. Investigation clicks were emitted randomly and considered monosyllabic [Table I and Fig. 2(k), bats=6, records=9]. They were similar to other click syllables.

5. Human interaction

All bats (males, females, and juveniles) produced these calls during interactions with human handlers.

Food clicks. Food clicks were only produced while the handler prepared food but were not accompanied by any specific behaviors. They were similar to other click syllables; however, they were louder and always emitted singly; hence, we did not perform temporal analyses [Fig. 2(l), bats=15, records=17].

Food solicitation calls. Bats emitted these calls during or immediately prior to feeding time. While producing these calls, bats approached the handler, sometimes hopping repeatedly and flapping their wings. Food solicitation calls were multisyllabic [Table I and Fig. 2(m), bats=8, records =18] and comprised of 2–11 (Table II) downward FM sweeps.

Alarm calls. These calls were associated with high levels of aggression and we only recorded these vocalizations from rabid bats or when wild bats were handled for the first time in captivity. Calling bats rose up on straightened elbows, jerked their bodies, and bared their teeth. Alarm calls were monosyllabic and emitted alone or in bouts [Table I and Fig.

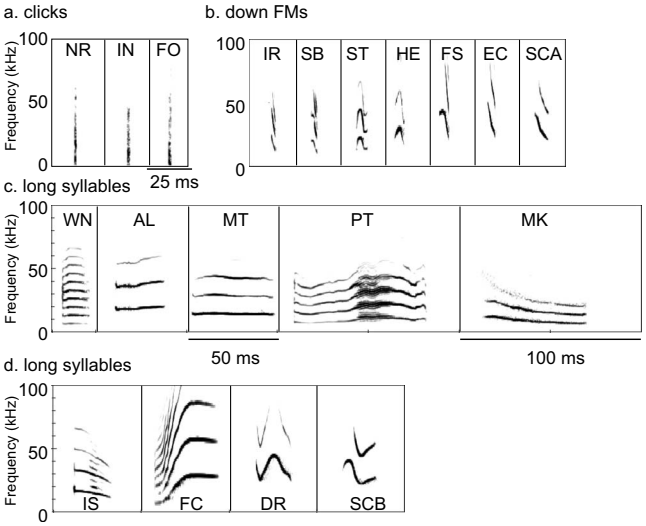


FIG. 4. Spectrograms of syllable types. (a) Clicks, NR=nose rubbing, IN=investigation, and FO=food click. (b) Brief downward frequency modulated syllables (down FMs), IR=irritation, SB=song buzz, ST=song trill, HE=herding, FS=food solicitation, EC=echolocation, and SCA=song chirp type A. [(c) and (d)] Long syllables, WN=warning, AL=alarm, MT=mounting, PT=protest, MK=marking, IS=isolation, FC=face rubbing, DR=directive, and SCB=song chirp type B.

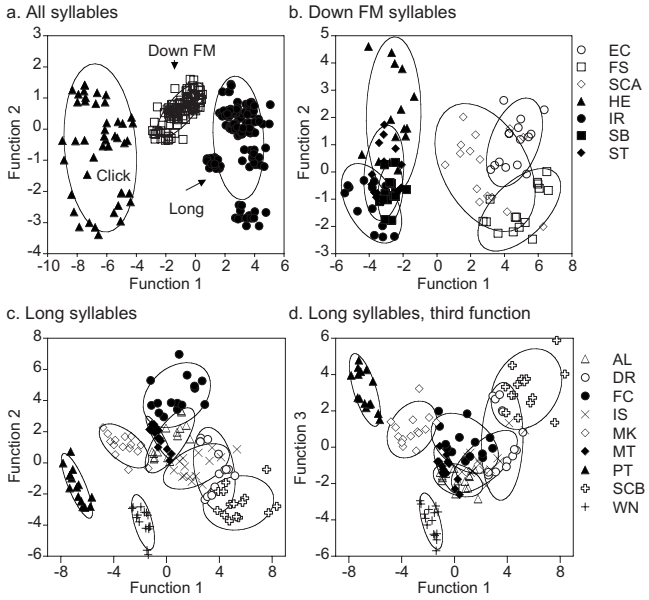


FIG. 5. Cannonical functions from discriminant function analyses. (a) All syllables fell into three distinct types. Function 1 was correlated with duration (r=0.99) and function 2 was correlated with peak frequency (r=0.99). The lines depict 80% normal bivariate density ellipses. (b) Down FM syllables. Function 1 was correlated with all variables. Function 2 was correlated with end and peak frequencies. We recognized two syllable subgroups: The first subgroup consisted of echolocation (EC), escape (ES), food solicitation (FS), and song chirp type A (SCA) syllables (open symbols) and the second subgroup consisted of irritation (IR), song buzz (SB), herding (HE), and song trill (ST) syllables (filled symbols). [(c) and (d)] Long syllables. Function 1 was correlated with frequency features, function 2 was most correlated with duration (c), and function 3 was most correlated with maximum frequency (d). AL=alarm, DR=maternal directive, FC=face rubbing, IS=isolation, MK=marking, MT=mounting, PT=protest, and SCB=song chirp type B. The lines depict 80% normal bivariate density ellipses.

2(n), bats=3, records=10]. Syllables were high amplitude and constant frequency (bandwidths less than 0.3 octaves).

6. Flight

In order to compare social vocalizations with echolocation signals, we recorded calls from three bats as they flew inside the barn.

Echolocation calls. Echolocation calls were monosyllabic and often emitted continuously [Table I and Fig. 2(o), bats=3, records=6]. They were short downward FM sweeps.

B. Syllable acoustics and temporal patterns

In this section, we (1) examine the acoustic features of syllables from each call, (2) determine whether each of the different calls is composed of a unique syllable, and (3) examine temporal patterns and call composition of calls that share common syllables. We examined the acoustic structure of 19 types of syllables. One for each call type and four types of syllables from song: type A and type B syllables in chirps, trill syllables, and buzz syllables.

First, we found that syllables fit into three distinct acoustic groups: (1) clicks, (2) down FMs, and (3) long syllables (Fig. 4). Discriminant function analysis confirmed these categories with 100% correct classification [Wilk's lambda =0.067, df=4,562, $P<0.0001$, Fig. 5(a)]. Duration was the most distinctive feature of these groups. Clicks were less than 1 ms, down FMs ranged between 1 and 8 ms, and long syllables ranged between 8 and 106 ms. With the exception of song, calls were composed of syllables from only one major acoustic group. For example, the only syllables in nose-rubbing calls were clicks while the syllables in irritation and food solicitation calls were all down FMs (Figs. 2 and 4). Since calls composed of syllables from different acoustic groups, clicks, down FMs, or long syllables should be highly distinguishable, we determined whether calls in each acoustic group can be distinguished by other features, specifically by syllable acoustics, temporal patterns, or call composition.

1. Clicks

The four acoustic features we measured (dur, f_{min}, f_{max}, and peak) did not differ between click calls (MANOVA; Wilk's lambda=0.15, df=8,6; P=0.43; mean ± SE; dur: 0.34 ± 0.04 ms, f_{min}=4.2 ± 0.8 kHz, f_{max}=32.3 ± 2.4 kHz, peak=15.0 ± 1.9). Therefore, the various calls composed solely of clicks are acoustically indistinguishable by the clicks themselves. However, the distribution of syllable emissions did differ between click calls. Anticipation clicks were only emitted singly, question clicks were emitted randomly, and nose-rubbing calls were the only multisyllabic calls where clicks were always emitted in distinct groups.

2. Down FMs

We examined four acoustic features of syllables from down FM calls (Table III). We found that syllables differed among calls (MANOVA: Wilk's lambda=0.001, df=24,39, $P<0.0001$) and bats (Wilk's lambda=0.08, df=56,317, $P<0.0001$). In order to examine which calls could be distinguished by syllable acoustics, we performed a discriminant function analysis. This analysis resulted in two canonical

TABLE III. Mean ± SE of down FM syllable measurements.

Syllable[a]	Code	Subgroup[b]	Temporal feature[c]	dur	f_{beg}	f_{end}	f_{peak}
Echo	EC	1	Mono	5.1 ± 0.2	52.1 ± 0.9	28.1 ± 0.9	35.9 ± 1.1
Food	FS	1	Multi	5.9 ± 0.3	49.9 ± 1.4	22.0 ± 1.0	27.9 ± 1.1
Song A	SCA	1	Context	5.0 ± 0.3	43.9 ± 1.5	21.0 ± 0.4	35.0 ± 2.3
Herding	HE	2	Context	3.2 ± 0.2	25.5 ± 1.0	21.6 ± 1.4	26.8 ± 1.1
Irritation	IR	2	ISI	2.3 ± 0.2	18.0 ± 1.0	11.2 ± 0.6	13.0 ± 0.4
Buzz	SB	2	Number	2.9 ± 0.1	22.2 ± 0.5	12.4 ± 0.3	19.5 ± 1.1
Trill	ST	2	Context	3.4 ± 0.3	20.7 ± 0.8	13.6 ± 0.9	22.2 ± 0.7

Food=food solicitation, song A=song chirp type A syllable, and buzz and trill are from song.
Discriminant function analysis revealed two acoustic subgroups.
Temporal features that differentiated calls from the same subgroup. Call distribution (mono- or multisyllabic), acoustic context (context), intersyllable interval (ISI), and number of syllables (number).

functions with eigenvalues greater than 1, which together explained 96% of the variation in the four acoustic variables [Fig. 5(b)]. Examination of the distribution of syllables on the first two canonical functions revealed two clusters or "subgroups" of down FM syllables. Subgroup 1 contained echolocation (EC), food solicitation (FS), and song type A syllables (SCA) and subgroup 2 contained herding (HE), irritation (IR), song buzz (SB), and song trill (ST) syllables. Subgroup 1 syllables were generally longer and higher in frequency than subgroup 2 syllables (Table III). Cross-validation resulted in 30 misclassifications (28%), all of which occurred within subgroups ($N=105$). This analysis suggests that syllables from different subgroups should be distinguishable whereas syllables from the same subgroup may be confused with each other.

Next, we examined whether temporal patterns or call composition differed between calls that were composed of syllables from the same down FM subgroup (Table III). For subgroup 1, food solicitation calls were the only multisyllabic calls and hence could be differentiated from echolocation calls and song type A syllables. Type A syllables and echolocation calls differed only by acoustic context; type A syllables were embedded in complex song phrases.

The four vocalizations in subgroup 2, herding calls, irritation calls, song buzz phrases, and song trill phrases, all had similar temporal emission distributions. Herding and irritation calls were both multisyllabic calls (Fig. 2) and we considered song buzz and trill phrases multisyllabic "calls" as well since they were always produced in distinct groups (Fig. 3). We tested whether these calls or phrases differed by the interval between syllables or the number of syllables per call/phrase (Table III). We found that irritation calls had longer intersyllable intervals ($F_{3,36}=34.0$, $P<0.001$, Tukey test $Q=2.69$, $\alpha=0.05$, Table II) than the other calls/phrases and song buzzes had more syllables ($F_{3,39}=83.3$, $P<0.0001$, $\log_{10}$-transformed data; Tukey test, $Q=2.65$, $\alpha=0.05$; Tables II and III). Herding calls and song trills did not differ in either the number of syllables or intersyllable intervals and may only be distinguishable by whether or not they are embedded within songs.

3. Long syllables

We examined six acoustic features of long call syllables (Table IV). Long syllable acoustics differed among calls (MANOVA, Wilk's lambda=0.000 04, $df=48,68$, P

TABLE IV. Mean ± SE of long syllable measurements (ms and kHz).

Syllable[a]	Code	Subgroup[b]	Temporal feature[c]	dur	f_{beg}	f_{end}	f_{min}	f_{max}	f_{peak}
Marking	MK	None	⋯	58.6 ± 3.4	9.1 ± 0.7	6.7 ± 1.0	6.5 ± 0.3	9.8 ± 1.5	7.9 ± 0.5
Protest	PT	None	⋯	60.0 ± 5.1	9.2 ± 0.6	9.7 ± 0.8	8.1 ± 0.6	11.0 ± 0.8	1.9 ± 0.1
Warning	WN	None	⋯	13.4 ± 0.8	7.2 ± 0.3	6.6 ± 0.2	6.3 ± 0.2	7.3 ± 0.2	7.0 ± 0.2
Directive	DR	1	Multi	21.4 ± 0.6	25.1 ± 1.4	18.6 ± 2.1	15.5 ± 1.4	29.9 ± 2.4	23.0 ± 0.4
Isolation	IS	1	Mono	23.3 ± 0.5	22.6 ± 1.6	15.3 ± 1.4	15.1 ± 1.4	22.8 ± 1.5	19.4 ± 1.3
Song B	SCB	1	Context	17.4 ± 0.5	39.6 ± 1.4	27.7 ± 1.1	25.9 ± 1.4	48.5 ± 2.5	29.1 ± 2.4
Alarm	AL	2	Rate	29.4 ± 1.2	18.8 ± 1.1	18.8 ± 1.1	17.8 ± 1.1	20.0 ± 1.0	18.5 ± 1.2
Face rub	FC	2	Multi	50.5 ± 3.9	13.5 ± 1.9	23.4 ± 1.0	12.4 ± 1.7	25.1 ± 1.0	23.7 ± 1.3
Mounting	MT	2	Rate	37.5 ± 2.0	15.3 ± 0.5	15.7 ± 0.4	14.1 ± 0.4	16.8 ± 0.3	15.4 ± 0.2

Directive=maternal directive, isolation=infant isolation, and song B=song chirp type B syllable.
Discriminant function analysis revealed two clusters of calls. Subgroup 1 syllables were longer and had some frequency features that were higher than subgroup 2. "None" refers to syllables that did not overlap with syllables from any other calls.
Temporal features that differentiated calls from the same subgroup. Call distribution (mono- or multisyllabic), acoustic context (context), and call rate (rate).

<0.0001) and bats (Wilk's lambda=0.0009, $df=108,597$, $P<0.0001$). Canonical discrimination resulted in three canonical functions with eigenvalues greater than 1 that together explained 94% of the variation in the six acoustic variables [Figs. 5(c) and 5(d)]. Warning (WN), protest (PT), and marking (MK) syllables did not overlap with syllables from other calls on one or more canonical functions [Figs. 5(c) and 5(d)]. These syllables generally had fundamental frequencies that were lower than the other calls (Table IV). A cross-validation procedure correctly classified all of these syllables. Thus, calls composed of these syllables (WN, PT, and MK) are likely distinguishable from other calls.

The remaining syllables were involved in 26 misclassifications (19% of 135 syllables). All errors occurred within two clusters or subgroups of calls. Subgroup 1 contained maternal directive (DR), song type B (SCB), and isolation (IS) syllables and subgroup 2 included alarm (AL), face-rubbing (FC), and mounting (MT) syllables. Syllables from subgroup 1 were generally briefer and had higher beginning frequencies than subgroup 2 (Table IV). This analysis suggests that calls composed of syllables from different subgroups should be distinguishable, whereas calls composed of syllables from the same subgroup may be confused with each other.

Next, we tested whether calls composed of syllables from the same subgroup differed in temporal emission patterns or call composition (Table IV). Maternal directive calls, isolation calls, and song type B syllables, which comprised subgroup 1, all differed by their emission characteristics. Maternal directive and isolation calls had different temporal emission patterns since maternal directives are multisyllabic and isolation calls are monosyllabic. Song type B syllables differed from the others by their acoustic context because they were embedded within complex phrases in song.

All calls in subgroup 2 (alarm, face-rubbing, and mounting calls) differed in either temporal emission patterns or call composition (Table IV). First, face-rubbing calls differed from the others because they were the only multisyllabic calls in this subgroup. Both alarm and mounting calls were monosyllabic; however, alarm calls were emitted at much lower rates than mounting calls ($F_{1,32}=23.5$, $P<0.0001$, alarm calls$=1.0\pm0.2$ calls/s, mounting calls$=5.5\pm0.6$ calls/s).

IV. DISCUSSION

Here, we showed that Mexican free-tailed bats use a rich repertoire of signals for vocal communication. We identified 16 types of vocalizations by the behavioral contexts in which they were emitted. We used quantitative analyses of temporal emissions to determine intercall/intracall boundaries and to define calls as monosyllabic or multisyllabic. We measured spectral and temporal features for syllables from each call and found that many, but not all calls, could be distinguished by the acoustics of their syllables. For those calls that did not have distinctive syllables, most had different temporal emission patterns. Some syllables and phrases were shared between calls and male song. Below, we discuss the implications of each of the major findings and how they relate to previous studies on vocal communication.

A. Vocal diversity

Our results suggest that vocalizations emitted by *T. brasiliensis* are as diverse as other bat species (i.e., *Pteronotus parnelli*, Kanwal *et al.*, 1994; *Rhinolophus ferrumequinum*, Ma *et al.*, 2006; *Saccopteryx bilineata*, Davidson and Wilkinson, 2002). We found that *T. brasiliensis* use diverse syllable types with broad spectral, temporal, and spectrotemporal features. Fundamental frequencies ranged from 1.5 to 35 kHz, overall frequencies from 1 to 95 kHz, and durations from 1 to 100 ms. We also observed diverse spectrotemporal patterns. Syllable bandwidths varied from relatively constant frequencies to up to nine octaves and spectrotemporal shapes included downward, upward, sinusoidal, and broadband (noisy) components. Although we observed a highly diverse repertoire, our study does not report the entire spectrum of vocal diversity in this species since we only used vocalizations if they were unambiguously correlated with recognizable behaviors and observed in multiple individuals.

B. Common syllables

One of our major findings was that syllables were not necessarily unique to a particular call, but rather the same syllables were used in different calls or song. One possible benefit of using the same syllables in different calls is that it allows for an expansion of the vocal repertoire without producing new syllables. This is well illustrated by the ubiquitous use of down FM syllables. For example, the down FM syllables in irritation calls were acoustically indistinguishable from those in the buzz phrases of song.

Not only were syllables shared among different social vocalizations, we also found that FM pulses emitted for echolocation were used in social communication, and were even embedded within song (song type A syllables). Recent research has shown that echolocation buzzes emitted while bats are foraging are also nearly indistinguishable from buzzes from songs (Schwartz *et al.*, 2007). The idea that echolocation calls may serve communicative functions is not new (reviewed in Altringham and Fenton, 2003). Echolocation calls have been shown to contain information on identity (Masters *et al.*, 1995), age (Masters *et al.*, 1995), sex (Kazial and Masters, 2004), or social group (Pearl and Fenton, 1996) and in at least one study bats also responded to this information (Kazial and Masters, 2004). Echolocation calls can also be emitted intermittently with social calls or may be modified to become social calls (Fenton, 2003). In *T. brasiliensis*, however, the use of echolocation pulses is different in that they are used while performing courtship displays with no echolocation function and are embedded into highly stereotyped social communication signals.

C. Temporal patterns

The most salient differences between some calls were their temporal emission patterns rather than syllable acous-

tics. Clicks are a good example. Although questioning clicks, anticipation clicks, and nose-rubbing calls are emitted in very different behavioral contexts and likely convey different information, they do not differ in the acoustic features of the click syllables themselves. These calls have different syllable distributions: Questioning clicks were emitted randomly, anticipation clicks were emitted singly, and nose-rubbing calls were always emitted in distinctive groups. Interestingly, temporal emission patterns are also important in sperm whales (Weilgart and Whitehead, 1997), one of the only mammals other than bats (Schmidt-French *et al.*, 2006) that use clicks for social communication. Sperm whales emit codas, which are highly stereotyped series of clicks (Rendell and Whitehead, 2005). Codas vary between social groups in both the number of clicks and the pattern of intervals between clicks. Although we use clicks as an example, temporal emissions patterns are also important for differentiating between long calls and between down FM calls.

Calls not only varied in the distribution of temporal emissions (i.e., whether or not syllables were clumped) but also in emission features. For example, the emission rate (calls/s) varied greatly between calls with similar syllable acoustics (e.g., mounting and alarm calls). On the other hand, the most salient difference for some multisyllabic vocalizations, such as calls and phrases composed of low frequency down FM syllables, was the number of syllables per call. Although syllable number may be rather important and stereotyped in some vocalizations, in other calls, such as face-rubbing and nose-rubbing calls, syllable number was more variable. Thus, in these calls, syllable number may not be a reliable feature even if the calls are always produced in a multisyllabic fashion.

D. Call composition

We found that multiple components of song were similar to other calls. For example, type A syllables in song were similar to echolocation pulses and trill phrases were similar to herding calls. Thus, both individual elements (syllables) and multiple elements (multisyllabic calls) were embedded into more complex vocalizations. One implication of this finding is that information may be conveyed by the acoustic context of syllables or the way in which they are combined.

Combining calls or notes to form new sequences has been observed in other taxa (Cleveland and Snowdon, 1982; Robinson, 1984; Hailman and Ficken, 1986; Crockford and Boesch, 2005; Arnold and Zuberbühler, 2006a). In some birds and primates, higher order combinations or the presence of particular elements elicit different responses from receivers and thus appear to affect signal meaning (Freeberg and Lucas, 2002; Clucas *et al.*, 2004; Arnold and Zuberbühler, 2006b). Future research on how syllables are combined and perceived may determine whether this is also the case in bats.

E. Determining call boundaries

In this study, we used a quantitative analysis of temporal patterns to determine whether calls were mono- or multisyllabic and to determine inter-and intracall boundaries. Most studies have either not addressed temporal emission patterns or relied on a qualitative approach for determining call (or bout) boundaries over time. For example in bats, bouts of vocalizations (termed "calls," "syllable trains," or "phrases") have been identified as sequences of syllables where the interval between syllables is less than the duration of the preceding syllable, less than the duration of the two syllables surrounding the interval, or less than the mean length of the two neighboring syllables (Kanwal *et al.*, 1994; Behr and von Helversen, 2004; Ma *et al.*, 2006). These methods are inconsistent between studies and impossible to implement across vocalizations with a broad range of durations (in this study, 1–100 ms) or intervals between syllables (in this study, 3–100 ms). Furthermore, since these thresholds are set by the researchers, they often do not reflect salient features of the vocalizations themselves. For example, using one of the criteria mentioned above, maternal directives would be considered monosyllabic (average duration =21 ms, Table III; average intersyllable interval within calls=60 ms, Table II) even though these vocalizations are always produced in distinct groups (Fig. 1).

An important aspect of our examination of temporal emission patterns is that they provide working hypotheses for vocal perception. Our results suggest that the meaning conveyed by calls may also be determined in part by the temporal emissions of syllables. In primates, research has shown that for some vocalizations, multiple syllables are the units of perception (Hauser *et al.*, 1998; Ghazanfar *et al.*, 2001) rather than single syllables. In *T. brasiliensis*, we hypothesize that for multisyllabic calls that are always emitted in groups and never alone, the groups of syllables are likely the units of perception rather than individual syllables. Perceptual research has also indicated that temporal emission features, such as the number of syllables per call and the interval between syllables, are important for call perception in primates (Hauser *et al.*, 1998; Ghazanfar *et al.*, 2002) and may also be the case in bats.

ACKNOWLEDGMENTS

This research was funded by the National Institutes of Health (Grant No. DC 20068). We thank Amanda Lollar for her input, Carl Resler and Sari Andoni for technical assistance, and Rachel Page and Josh Gittelman for helpful comments on the manuscript.

Altringham, J. D., and Fenton, M. B. (**2003**). "Sensory ecology and communication in the chiroptera," in *Bat Ecology*, edited by T. H. Kunz and M. B. Fenton (The University of Chicago Press, Chicago, IL), pp. 90–127.

Arnold, K., and Zuberbühler, K. (**2006a**). "The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*," Anim. Behav. **72**, 643–653.

Arnold, K., and Zuberbühler, K. (**2006b**). "Semantic combinations in primate calls," Nature (London) **441**, 303.

Balcombe, J. P. (**1990**). "Vocal recognition of pups by mother Mexican free-tailed bats, *Tadarida brasiliensis mexicana*," Anim. Behav. **39**, 960–966.

Balcombe, J. P., and McCracken, G. F. (**1992**). "Vocal recognition in Mexican free-tailed bats—do pups recognize mothers?," Anim. Behav. **43**, 79–87.

Barclay, R. M. R., Fenton, M. B., and Thomas, D. W. (**1979**). "Social-behavior of the little brown bat, *Myotis lucifugus*. II. Vocal communication," Behav. Ecol. Sociobiol. **6**, 137–146.

Barlow, K. E., and Jones, G. (**1997**). "Differences in songflight calls and social calls between two phonic types of the vespertilionid bat Pipistrellus pipistrellus," J. Zool. **241**, 315–324.
Behr, O., and von Helversen, O. (**2004**). "Bat serenades-complex courtship songs of the sac-winged bat (*Saccopteryx bilineata*)," Behav. Ecol. Sociobiol. **56**, 106–115.
Bohn, K. M., Wilkinson, G. S., and Moss, C. F. (**2007**). "Discrimination of infant isolation calls by greater spear-nosed bats, *Phyllostomus hastatus*," Anim. Behav. **73**, 423–432.
Boughman, J. W. (**1997**). "Greater spear-nosed bats give group-distinctive calls," Behav. Ecol. Sociobiol. **40**, 61–70.
Boughman, J. W. (**1998**). "Vocal learning by greater spear-nosed bats," Proc. R. Soc. London, Ser. B **265**, 227–233.
Catchpole, C. K., and Slater, P. J. B. (**1995**). *Bird Song: Biological Themes and Variations* (Cambridge University Press, Cambridge).
Cheney, D. L., and Seyfarth, R. M. (**1980**). "Vocal recognition in free-ranging vervet monkeys," Anim. Behav. **28**, 362–367.
Cleveland, J., and Snowdon, C. T. (**1982**). "The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*)," Z. Tierpsychol **58**, 231–270.
Clucas, B. A., Freeberg, T. M., and Lucas, J. R. (**2004**). "Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*)," Behav. Ecol. Sociobiol. **57**, 187–196.
Crockford, C., and Boesch, C. (**2005**). "Call combinations in wild chimpanzees," Behaviour **142**, 397–421.
Davidson, S. M., and Wilkinson, G. S. (**2002**). "Geographic and individual variation in vocalizations by male *Saccopteryx bilineata* (Chiroptera: Emballonuridae)," J. Mammal. **83**, 526–535.
Esser, K.-H. (**1994**). "Audio-vocal learning in a nonhuman mammal—the lesser spear-nosed bat *Phyllostomus discolor*," NeuroReport **5**, 1718–1720.
Esser, K.-H., and Schmidt, U. (**1989**). "Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllosomidae)—Evidence for acoustic learning," Ethology **82**, 156–168.
Fenton, M. B. (**2003**). "Eavesdropping on the echolocation and social calls of bats," Mammal Rev. **33**, 193–204.
Freeberg, T. M., and Lucas, J. R. (**2002**). "Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*," Anim. Behav. **63**, 837–845.
French, B., and Lollar, A. L. (**1998**). "Observations on the reproductive behavior of captive *Tadarida brasiliensis mexicana* (Chiroptera: Molossidae)," South. Nat. **43**, 484–490.
Gelfand, D. L., and McCracken, G. F. (**1986**). "Individual variation in the isolation calls of Mexican free-tailed bat pups (*Tadarida brasiliensis mexicana*)," Anim. Behav. **34**, 1078–1086.
Ghazanfar, A. A., Flombaum, J. I., Miller, C. T., and Hauser, M. D. (**2001**). "The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): Playback experiments with long calls," J. Comp. Physiol., A **187**, 27–35.
Ghazanfar, A. A., Smith-Rohrberg, D., Pollen, A. A., and Hauser, M. D. (**2002**). "Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins," Anim. Behav. **64**, 427–438.
Gould, E., Woolf, N. K., and Turner, D. C. (**1973**). "Double-note communication calls in bats—Occurrence in 3 families," J. Mammal. **54**, 998–1001.
Gouzoules, H., and Gouzoules, S. (**1990**). "Matrilineal signatures in the recruitment screams of pigtail macaques, *Macaca nemestrina*," Anim. Behav. **115**, 327–347.
Hailman, J. P., and Ficken, M. S. (**1986**). "Combinational animal communication with computable syntax: Chick-a-dee calling qualifies as 'language' by structural linguistics," Anim. Behav. **34**, 1899–1901.
Hauser, M. D. (**1991**). "Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations," Ethology **89**, 29–46.
Hauser, M. D. (**1998**). "Functional referents and acoustic similarity: Field playback experiments with rhesus monkeys," Anim. Behav. **55**, 1647–1658.
Hauser, M. D., Agnetta, B., and Perez, C. (**1998**). "Orienting asymmetries in rhesus monkeys: The effect of time-domain changes on acoustic perception," Anim. Behav. **56**, 41–47.
Kanwal, J. S., Matsumura, S., Ohlemiller, K., and Suga, N. (**1994**). "Analysis of acoustic elements and syntax in communication sounds emitted by moustached bats," J. Acoust. Soc. Am. **96**, 1229–1254.
Kazial, K. A., and Masters, W. M. (**2004**). "Female big brown bats, *Eptesicus fuscus*, recognize sex from a caller's echolocation signals," Anim. Behav. **67**, 855–863.
Kroodsma, D. E. (**1977**). "A re-evaluation of song development in the song sparrow," Anim. Behav. **25**, 390–399.
Lollar, A., and Schmidt-French, B. (**2002**). *Captive Care and Medical Reference for the Rehabilitation of Insectivorous Bats* (Bat World Publication, Mineral Wells, WL).
Ma, J., Kobayasi, K., Zhang, S., and Metzner, W. (**2006**). "Vocal communication in adult greater horseshoe bats, *Rhinolophus ferrumequinum*," J. Comp. Physiol., A **192**, 535–550.
Marler, P. (**2004**). "Bird calls: Their potential for behavioral neurobiology," Ann. N.Y. Acad. Sci. **1016**, 31–44.
Marler, P., and Slabbekoorn, H. (**2004**). *Nature's Music, The Science of Birdsong* (Elsevier, Amsterdam).
Masters, W. M., Raver, K. A. S., and Kazial, K. A. (**1995**). "Sonar signals of big brown bats *Eptesicus fuscus*, contain information about individual identity, age and family affiliation," Anim. Behav. **50**, 1243–1260.
Melendez, K. V., Jones, D. L., and Feng, A. S. (**2006**). "Classification of communication signals of the little brown bat," J. Acoust. Soc. Am. **120**, 1095–1102.
Pearl, D. L., and Fenton, M. B. (**1996**). "Can echolocation calls provide information about group identity in the little brown bat (*Myotis lucifugus*)," Can. J. Zool. **74**, 2184–2192.
Pfalzer, G., and Kusch, J. (**2003**). "Structure and variability of bat social calls: Implications for specificity and individual recognition," J. Zool. **261**, 21–33.
Porter, F. L. (**1979**). "Social behavior in the leaf-nosed bat, *Carollia perspicillata. II*. Social communication," Z. Tierpsychol **50**, 1–8.
Range, F., and Fischer, J. (**2004**). "Vocal repertoire of sooty mangabeys (*Cercocebus torguatus atys*) in the Tai National Park," Ethology **110**, 301–321.
Rendell, L., and Whitehead, H. (**2005**). "Spatial and temporal variation in sperm whale coda vocalizations: Stable usage and local dialects," Anim. Behav. **70**, 191–198.
Robinson, J. G. (**1984**). "Syntactic structures in the vocalizations of wedge capped capuchin monkeys Cebus nigrivittatus," Behaviour **90**, 46–79.
Russ, J. M., Jones, G., Mackie, I. J., and Racey, P. A. (**2004**). "Interspecific responses to distress calls in bats (Chiroptera: Vespertilionidae): A function for convergence in call design?," Anim. Behav. **67**, 1005–1014.
Russ, J. M., Racey, P. A., and Jones, G. (**1998**). "Intraspecific responses to distress calls of the pipistrelle bat, *Pipistrellus pipistrellus*," Anim. Behav. **55**, 705–713.
Ryan, M. J. (**2001**). *Anuran Communication* (Smithsonian Institution Press, Washington, DC).
Schmidt-French, B., Gillam, E., and Fenton, M. B. (**2006**). "Vocalizations emitted during mother/young interactions by captive eastern red bats *Lasiurus borealis* (Chiroptera: Vespertilionidae)," Acta Chiropt. **8**, 477–484.
Schwartz, C., Tressler, J., Keller, H., Vanzant, M., Ezell, S., and Smotherman, M. (**2007**). "The tiny difference between foraging and communication buzzes utered by the Mexican free-tailed bat, *Tadarida brasiliensis*," J. Comp. Physiol., A **193**, 853–863.
Searby, A., Jouventin, P., and Aubin, T. (**2004**). "Acoustic recognition in macaroni penguins: An original signature system," Anim. Behav. **67**, 615–625.
Seddon, N., Tobias, J. A., and Alvarez, A. (**2002**). "Vocal communication in the pale-winged trumpeter (*Psophia leucoptera*): Repertoire, context and functional reference," Behaviour **139**, 1331–1359.
Sibly, R. M., Nott, H. M. R., and Fletcher, D. J. (**1990**). "Splitting behaviour into bouts," Anim. Behav. **39**, 63–69.
Todt, D., Goedeking, P., and Symmes, D. (**1988**). *Primate Vocal Communication* (Springer-Verlag, New York).
Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior in dolphins and whales," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 156–224.
Weilgart, L., and Whitehead, H. (**1997**). "Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales," Behav. Ecol. Sociobiol. **40**, 277–285.
Wilkinson, G. S., and Boughman, J. W. (**1998**). "Social calls coordinate foraging in greater spear-nosed bats," Anim. Behav. **55**, 337–350.
Zar, J. H. (**1984**). *Biostatistical Analysis*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ).

Modeling the sound levels produced by bubble release of individual herring

Thomas R. Hahn and Gary Thomas
University of Miami, RSMAS, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 28 May 2007; revised 30 May 2008; accepted 2 June 2008)

Herring (*Clupea pallasii and C. harengus*) are known to release gas from their swim bladder to assist a number of complex behaviors, such as buoyancy adjustments and predator avoidance. The noise associated with the release has recently been reported in the literature and related to oscillating bubbles. Average source levels (SLs) of 73 dB with regard to μPa rms reference 1 m have been reported for bubbles produced by herring in the laboratory. A model is provided for predicting the SL in terms of the gas flow rate from the swim bladder into the bubbles. Based on these laboratory conditions, an inversion yields a rate of 0.9 (0.3–3.2) ml/min. Furthermore, the model predicts an acoustic SL of 89 (79–99) dB with regard to μPa rms reference 1 m for pulses emitted by herring in a natural shallow water environment at unknown distance corresponding to a flow rate of 2.5 ml/min. An analysis of published acoustic data suggests that herring is capable of controlling the gas flow and the corresponding acoustic levels over a wide range according to different behavioral needs. The proposed model allows an extrapolation of the laboratory results to situations that are relevant for bubble release of herring schools in the ocean.
 [DOI: 10.1121/1.2951593]

PACS number(s): 43.80.Ka, 43.30.Nb, 43.30.Sf [KGF] Pages: 1849–1857

I. INTRODUCTION

In many underwater environments, vision is severely limited in range; hence, acoustics is the logical means for animals to detect predators, prey, and each other as well as to navigate and communicate. Despite this importance, the generation and use of underwater sound by fishes are still poorly understood. While primary interests for herring have been in using active acoustics to measure biomass,[1] we have been exploring the use of passive acoustic techniques to improve the efficiency of surveys.

One aspect of the herring behavior that has fascinated us is the release of gas by adult herring during their winter, diel vertical migration.[2] These migrating aggregations have been observed to release massive plumes of gas at dusk, which we believe to be *triggered* by the reduction of static pressure as the herring attempt to stop their upward diel vertical migration to shallow nighttime depths. If detectible, the associated noise could be a means to directly determine the presence or the absence, and entry or exit of herring to and from their preferred winter embayments. This would be a highly valuable addition to the wintertime acoustic assessments because it could reduce the time spent searching for fish schools, verify aerial observations, and also provide a valuable insight into the fish's migratory timing and predator-avoidance behaviors.[3] If the local acoustic environment and the individual and school aspects of gas release and sound production were well understood, one might hope to use sound intensities, spectra, phase correlations, and range information for coarsely estimating their size and abundance.

Furthermore, winter aggregations of herring are under heavy predation by marine mammals[4] and birds,[5] which is also believed to cause herring to release gas. The ability to detect this release of gas could provide valuable information on the intensity of predation on adult herring during the winter months.[3] In a subsequent paper, we hypothesize that the release of gas during various behavioral modes may indeed provide sufficient noise to allow for detection under favorable circumstances. First, to strengthen our argument, we need to develop a firmer theoretical basis to explain the phenomenon and to provide a theoretical source level (SL) estimate applicable to various situations. This is the purpose of this work.

In a recent publications, Wahlberg and Westerberg[6] investigated the sound produced by herring (*Clupea harengus*) while releasing gas in form of bubbles. These investigators made measurements of the associated noise production in cages, *in situ*, and under laboratory conditions to simulate vertical migration. In another publication, Wilson *et al.*[7] independently studied gas release of herring (*C. pallasii and C. harengus*) in large tanks to investigate group behavior. Much has been learned from these first efforts to study herring bubble release acoustically. However, a more detailed understanding of the acoustics of the gas release and its detailed underlying physical mechanisms is necessary for stock assessment purposes and to improve our understanding of herring behavior.

A general mechanism was proposed for the production of gas bubbles by herring.[6] Our paper provides a theoretical understanding of the measured SL from first physical principles and in terms of physiological features of the herring gas release. These are often notoriously difficult to predict. However, they are the most salient ingredient in predicting the acoustic emissions of herring *schools*. Additionally, agreement of predicted unique acoustic signatures with field observations will provide further support for the proposed sound production mechanism. We will address this issue in the subsequent paper.

Under laboratory conditions in a low-pressure chamber (LPC), Wahlberg and Westerberg[6] found a SL of about 73 dB with regard to μPa rms reference 1 m for the chirps originating from gas release of individual herring (*C. harengus*). While absolute SLs of oscillating bubbles in chambers are generally difficult to interpret,[8] these authors apparently took great care to avoid reverberations (the LPC was constructed of impedance matched material and placed into a sufficiently deep body of water). The chirps are found to contain a larger number of individual pulses, often at decreasing frequency. The authors plausibly associate this signal with gas bubbles that are formed during the gas release of the observed specimens. It is well known that bubbles, during their formation stage, are strong sources of sound.[9] This is because bubbles are either statically or dynamically out of equilibrium when born by entrainment, breakup, or injection. While reaching an equilibrium state, acoustic energy is released into the medium. The amount of this acoustic energy can be estimated from first principles and from independent physical parameters of the measured signal.

An outline of the theoretical approach is given in the first part of Sec. II. The presentation proceeds with the theoretical model for power spectra of herring chirps. A discussion of the results including a comparison with the observations of Wahlberg and Westerberg and those of Wilson *et al.* as well as an outlook toward further developments and applications are given in Sec. III.

II. METHODS

A. General approach

To compute the intensity of the sound generated by the gas release of individual fish, or fish schools containing a large number of fish, recorded at a particular point of observation $\vec{r}$, we first note that the resulting signal is the sum of many independent and randomly occurring discrete "events." The events in question are the pulses emitted from single oscillating bubbles, created by the gas releasing fish. Individually, the duration of these pulses is of order of milliseconds. The time span of acoustic activity stretches up to several seconds for individual fish[6] and may be of the order of some tens of minutes for large vertically migrating herring schools. The detailed mechanism of this acoustic activity will be emphasized below. For now, let $p_b(\vec{r},\vec{r}_b;t)$ be the wave form of an acoustic pulse received at $\vec{r}$ at time $t=0$ due to a single source event at $\vec{r}_b$. The entire signal at $\vec{r}$ is the superposition of all such individual pulses received at randomly occurring times t_k within the observation time T:

$$p(\vec{r},t)=\sum_{k=1}^{N} p_{b_k}(\vec{r},\vec{r}_{b_k};t-t_k). \tag{1}$$

For ease of computation, we decompose the individual pulses into their harmonic components $p_\omega(\vec{r},\vec{r}_b)$,

$$p_b(\vec{r},\vec{r}_b;t)=\int e^{-i\omega t}p_\omega(\vec{r},\vec{r}_b)d\omega, \tag{2}$$

which, in turn, are determined by acoustic Green's function of the environment $G_\omega(\vec{r},\vec{r}_b)$ and the source distribution. In our case, the bubbles are assumed to be point sources, which is an extremely good approximation for bubbles in monopole mode, which have geometrical dimensions that are small compared to the wavelength. Therefore, the pulse shapes are simply the product of Green's function, the source spectrum $\tilde{f}_\omega$, and the source strength s_b:

$$p_\omega(\vec{r},\vec{r}_b)=\int s_b\tilde{f}_\omega\delta(\vec{r}'-\vec{r}_b)G_\omega(\vec{r},\vec{r}')dV'=s_b\tilde{f}_\omega G_\omega(\vec{r},\vec{r}_b) \tag{3}$$

The split between $\tilde{f}_\omega$ and s_b is, at this stage, somewhat artificial but will be chosen such that the former is universal for bubbles of a given size, while the latter describes the level of excitation of an individual bubble. Green's function $G_\omega(\vec{r},\vec{r}_b)$ contains the influence of the sound channel on the received pulse shape. For sound propagation of a wave with angular frequency ω in homogeneous free space, Green's function is given by $G_\omega(\vec{r},\vec{r}_b)=1/(4\pi)\exp(i\omega|\vec{r}-\vec{r}_b|/c)/|\vec{r}-\vec{r}_b|$.

For random pulse trains, the arrival times for pulses originating in a volume element dV_b of the fish schools are Poisson distributed. Therefore, the time averaged *one-sided* power spectral density of the received signal due to pulses originating in dV_b is given by[10] (compared with this reference, we use a different definition of the Fourier transform pairs, leading to a numerical factor of $8\pi^2$ instead of 2 in the following equation for the power spectra)

$$d\langle S_\omega\rangle=8\pi^2\rho_t\overline{\langle s_b^2\rangle|\tilde{f}_\omega|^2}|G_\omega(\vec{r},\vec{r}_b)|^2dV_b, \tag{4}$$

and the total power spectral density by

$$\langle S_\omega\rangle=8\pi^2\overline{\langle s_b^2\rangle|\tilde{f}_\omega|^2}\int\rho_t|G_\omega(\vec{r},\vec{r}_b)|^2dV_b. \tag{5}$$

The integration extends over the entire volume of acoustic activity and ρ_t denotes the number of acoustic events per unit time and volume, and if integrated over the entire active volume $\int\rho_t dV_b=\lim_{T\to\infty}N/T$, we obtain the total number of pulses N per observation time T, the pulse rate. The indicated instructions $\langle\,\rangle$ and $\bar{}$ refer to the ensemble average of the source strengths and, if desired, a subsequent averaging over bubble size distributions, respectively. This is the most general framework. In the case of a single stationary fish in free space, Eq. (5) simply reduces to

$$\langle S_\omega\rangle=\frac{\dot{n}}{2r^2}\overline{\langle s_b^2\rangle|\tilde{f}_\omega|^2}, \tag{6}$$

$\dot{n}$ denoting the total pulse rate, and r the receiver distance from the fish.

Using this scheme, the power spectral density of a gas releasing fish can, in essence, be estimated from the average source strength and the bubble size distribution of acoustically active bubbles together with the overall event rate. The spectral shape $\tilde{f}_\omega$ of an individual pulse is determined by the dynamics of individual bubble oscillations and is well known if free field conditions are assumed[9] (the interpretation of acoustic emissions in reverberant environments is significantly more complex[8]). All of these parameters can be inferred from the established measurements, or estimated with

sufficient accuracy to verify the plausibility of the underlying model. As we will see in the next section, this can be achieved and allows the assessment of basic production features of the gas release from the measured strength of the emitted sound.

B. Individual pulses and power spectra

Before turning to the estimation of the overall spectral levels, it is necessary to reexamine the acoustic levels and the spectral decomposition of *individual* acoustic source events. Wahlberg and Westerberg recorded the signal associated with gas release in herring in a LPC, in cages, and *in situ*. They found bubbles that result from the venting of gas into the water column to be responsible for the emission of a series of pulses of diminishing frequency maxima (chirped pulses). These authors conclude that, at least under controlled conditions, bubbles do not necessarily detach from the herring when emitting sound, which explains the chirped nature of the emitted sound. This indicates that the produced chirps are not predominantly the result of bubble separation but rather of discontinuities of the gas flow into the growing bubble, or in the words of Wahlberg and Westerberg: "For each portion of air added, the attached bubble is given an impulse to oscillate and generate sound at its resonance frequency."

To put this suggested sound production mechanism into a more comprehensive perspective, it is useful to look at principal ways radial oscillations of bubbles can be set into motion:[11] (a) abrupt changes in internal bubble pressure, e.g., due to surface tension, curvature, or hydrostatic effects, (b) mode coupling of shape oscillations initiated by shape distortions or asymmetries in the flow field,[12,13] (c) initial inward or outward radial fluid momentum,[14] or (d), jet injection into the bubble due to the specific initial bubble deformations.[15] Depending on the particular way air bubbles are created, some or all of these mechanisms are invoked to produce acoustic emissions. For example, bubble fragmentation is thought to play a major role in sound production by air-entraining processes at water surfaces.[9,16–19] For this particular case, all mechanisms can, in principle, provide energy for sound production. For ambient noise generation by breaking waves at the sea surface, surface tension closure pressure is believed to be dominant.

For slow bubble injection from an orifice leading to millimeter-sized bubbles, breakup will not occur during or after detachment because the Weber number is insufficiently high. In this case, if the bubble does not detach from the orifice, only mechanism (c) can provide substantial initial energy for the bubble oscillation. Therefore, we can reasonably assume that, in our case, the principal way to excite radial bubble oscillations is the initial outward momentum as fluid is displaced by the growing bubble. This is not to say that there could not be other small-scale mechanisms involved in the sound production that are peculiar to the physiology of herring. As these are currently out of experimental reach, we will ignore these for the purpose of this manuscript.

It is noteworthy that, if the flow rate of air is high enough, the initial loading of the bubble due to the outward momentum of the fluid is dominant even if the bubble detaches.[14] Based on the analysis of Longuet-Higgins,[15] we estimate that for most of our conditions, initial oscillation amplitudes due to changing surface tension pressure, are of smaller or equal magnitude compared to the momentum loading effect. Contributions of shape oscillations to the total levels of detaching bubbles are small in the far field if mode coupling is ignored. This seems reasonable to assume, since, measured in terms of curvature deviation from the spherical shape, distortions of bubbles that are about to detach from a nozzle are only about 10% for the sizes considered. This is much smaller than for bubbles breaking up in shear flows.

The hypothesis that the initial momentum surplus is delivering the excitation of the bubbles is not investigated experimentally in this study, but predictions based on this assumption will be shown to be consistent with the body of the available data. While suggesting a reasonable overall model for the mechanism of sound production on which we base our study, Wahlberg and Westerberg did not theoretically explain the observed sound *levels*. This is attempted here because of its importance in estimating the noise levels of massive bubble release of *entire schools* as reflected by the occurrence of the source strength s_b and its spectrum $\tilde{f}_\omega$ in Eq. (5).

A bubble, attached to the gas releasing fish and set into motion by the mechanism elaborated above, exhibits a damped oscillation of its radius $a(t)$ about its equilibrium size a_0:

$$a(t) = \mathrm{Re}[a_0 + a_\varepsilon e^{-i(\tilde{\omega}_0 t + \psi)}]. \tag{7}$$

The initial amplitude of this oscillation is indicated by a_ε, which is small compared to a_0 but otherwise a free variable, determined by the inertia of the medium at the point in time when the oscillation is set into motion. The frequency of oscillation ω_0 and the damping β of the bubble, once the bubble is freely "ringing," are essentially determined by the bubble equilibrium size a_0 and are for reasons of simplicity combined here in the complex-numbered quantity $\tilde{\omega}_0$:

$$\tilde{\omega}_0 = \omega_0 - i\beta. \tag{8}$$

The radiated sound pressure due to this radial bubble motion is obtained from the acceleration at the bubble wall (omitting the indication of the real part and the phase)

$$\dot{a}(t) = -i\tilde{\omega}_0 a_\varepsilon e^{-i\tilde{\omega}_0 t} = U_\varepsilon e^{-i\tilde{\omega}_0 t},$$

$$\ddot{a}(t) = -\tilde{\omega}_0^2 a_\varepsilon e^{-i\tilde{\omega}_0 t} = \dot{U}_\varepsilon e^{-i\tilde{\omega}_0 t}, \tag{9}$$

where U_ε is the initial amplitude of the bubble wall velocity and $\dot{U}_\varepsilon$ that of the bubble wall acceleration. This determines the acoustic pressure fluctuations $p_b(a,t)$ at the "wet" side of the bubble wall and, following the assumption of strict monopole radiation, everywhere in the fluid:[9]

$$p_b(r,t) \simeq \rho \frac{a_0^2}{r} \ddot{a}. \tag{10}$$

Equation (10) assumes that at the moment the bubble starts ringing, its gas content is disconnected from the gas reservoir in the fish. This assumption is fully consistent with the mechanism proposed by Wahlberg and Westerberg for the generation of the observed pulsed chirps.

The spectral shape of an individual pulse can be determined from the Fourier transform of the bubble wall acceleration,

$$\ddot{a}_\omega = -\frac{a_\varepsilon}{2\pi} \frac{\omega_0[\beta(\beta - 2i\omega) + \omega_0^2]}{(\beta - i\omega)^2 + \omega_0^2}, \tag{11}$$

which can be computed by direct evaluation of the Fourier integral of the real part of Eq. (9). Together with Eq. (10), the spectral shape of an individual pulse is determined to be

$$|\tilde{f}|^2 = \frac{4\beta}{\pi(\omega_0^2 + 5\beta^2)} \frac{\beta^2(\beta^2 + 4\omega^2 + 2\omega_0^2) + \omega_0^4}{[\beta^2 + (\omega - \omega_0)^2][\beta^2 + (\omega + \omega_0)^2]}. \tag{12}$$

The normalization factor has been chosen such that the one-sided spectral integral of Eq. (12) is unity. This choice of normalization leads to the following definition of the source strength, which encodes the level of monopole excitation of an individual bubble:

$$\langle s_b^2 \rangle = \frac{\rho^2 \langle \dot{V}^2 \rangle}{16\pi} \left(\frac{\omega_0^2}{\beta} + 5\beta \right). \tag{13}$$

It should be mentioned that a similar analysis has been followed to estimate the SL of bubbles ejected from cylindrical needles by Nicholas *et al.*[14]

The detailed observations of Wahlberg and Westerberg allow an estimation of the source strength s_b. If caused by the outward moving fluid, the initial amplitude of oscillation, a_ε, is related to the effective flow rate $\dot{V}$ of gas from the swim bladder at the moment the flow gets interrupted and the bubble momentarily disconnected from the air reservoir in the herring:

$$a_\varepsilon = \frac{1}{\omega_0} \frac{\dot{V}}{4\pi a_0^2}. \tag{14}$$

Finally, combining all ingredients into a single equation, we obtain for the power spectra of individual stationary fish chirps:

$$\langle S_\omega \rangle = \frac{\rho^2 \dot{n}}{8\pi^2 r^2} \langle \dot{V}^2 \rangle \overline{\frac{\beta^2(\beta^2 + 4\omega^2 + 2\omega_0^2) + \omega_0^4}{[\beta^2 + (\omega - \omega_0)^2][\beta^2 + (\omega + \omega_0)^2]}}. \tag{15}$$

The total emitted power is obtained by integrating $\langle S_\omega \rangle$ over the entire range of frequencies. The resulting integral can readily be evaluated by summing up the pole contribution from the upper half complex ω-plane. The result is

$$\frac{1}{T} \int_0^T p(\vec{r},t)^2 dt = \int_0^\infty \langle S_\omega \rangle \frac{d\omega}{2\pi} = \frac{\rho^2 \dot{n}}{64\pi^2 r^2} \langle \dot{V}^2 \rangle \overline{\left(\frac{\omega_0^2}{\beta} + 5\beta \right)}. \tag{16}$$

The "bubble dynamics term" in parenthesis features the resonance frequency ω_0 and the damping constant β, both of which are well understood for bubbles exhibiting small oscillations in free space. In the domain of our interest, ω_0 is well approximated by Minnaert's[20] formula $\omega_0 \simeq (1/a_0)\sqrt{3\gamma_{\text{eff}} P_0/\rho}$. The effective polytropic coefficient, γ_{eff}, is about 1.3–1.4 for millimeter-sized bubbles at resonance. The hydrostatic pressure at the locus of the bubble is denoted by P_0. Generally, β contains contributions due to the thermal and viscous effects as well as sound radiation.[9] For bubbles of 0.1–1 mm radius, the damping at the natural resonance frequency is mostly due to the thermal and radiation effects. A 1 mm radius bubble at its (sea-surface) resonance frequency has damping coefficients β=391 s^{-1} and 191 s^{-1} in the LPC at 0.2 bar. The corresponding values for 0.5 (and 1.5) mm radius bubbles are β=971 s^{-1} (231 s^{-1}) and β=500 s^{-1} (109 s^{-1}) at the LPC conditions.

We should note that the effective rms flow rate $\langle \dot{V}^2 \rangle^{1/2}$, introduced in Eq. (14), is not necessarily identical to the true volume flux out of the bladder. The underlying assumption made here is that the sudden change in the flow rate is of the same order of magnitude as the average volume flux. Such a situation would occur, for example, if a continuous stream of gas were interrupted *briefly* at the point in time when bubble oscillations are observed. Under this assumption, both quantities will be of similar magnitude.

The remaining issue to be addressed is the influence of the size distribution of the acoustically active bubbles on the average emitted power, Eq. (16). From the last pulses in a chirp, Wahlberg and Westerberg estimated the size distribution of the growing bubbles at their *final* stage. At first, this seems insufficient information to obtain the distribution taking into account *all* growth stages of the bubbles. However, this is not the case. Assuming that the incremental volume increase is small compared to the final bubble volume (or the final stage has statistically been reached in many substages), the *full* distribution of active bubble sizes a is independent from the statistics of the incremental volume additions and can directly be estimated from the distribution of the final-stage bubble sizes a_f. Two situations can be considered: first, if the signal considered is due to a larger number of released final-stage bubbles, and their distribution is known. Here, the upper half of this distribution is important, since it contains more events than the lower one and can even compensate for all the growing stages. Second, if the signal contains only pulses due to one growing bubble with definite finite size. In this case, clearly, the mean active bubble size will be lower than the final-stage size.

For the first case, we find that the mean of the distribution is $\bar{a} = (3M_4)/(4M_3)$, where M_n is the nth moment of the final-stage probability density function (PDF) $p(a_f)$. The same line of thinking leads to $\overline{a^2} = (3M_5)/(5M_3)$. This can be derived directly from the observation that the PDF in (a_f, a)-space is proportional to $4\pi a^2$, with many more events at the larger-size stages:

$$p(a_f, a) \propto 4\pi a^2 \Theta(a_f - a) p(a_f). \tag{17}$$

The Heaviside step function, $\Theta(a_f - a)$, ensures that bubbles do not grow beyond their final-stage sizes. For normally dis-

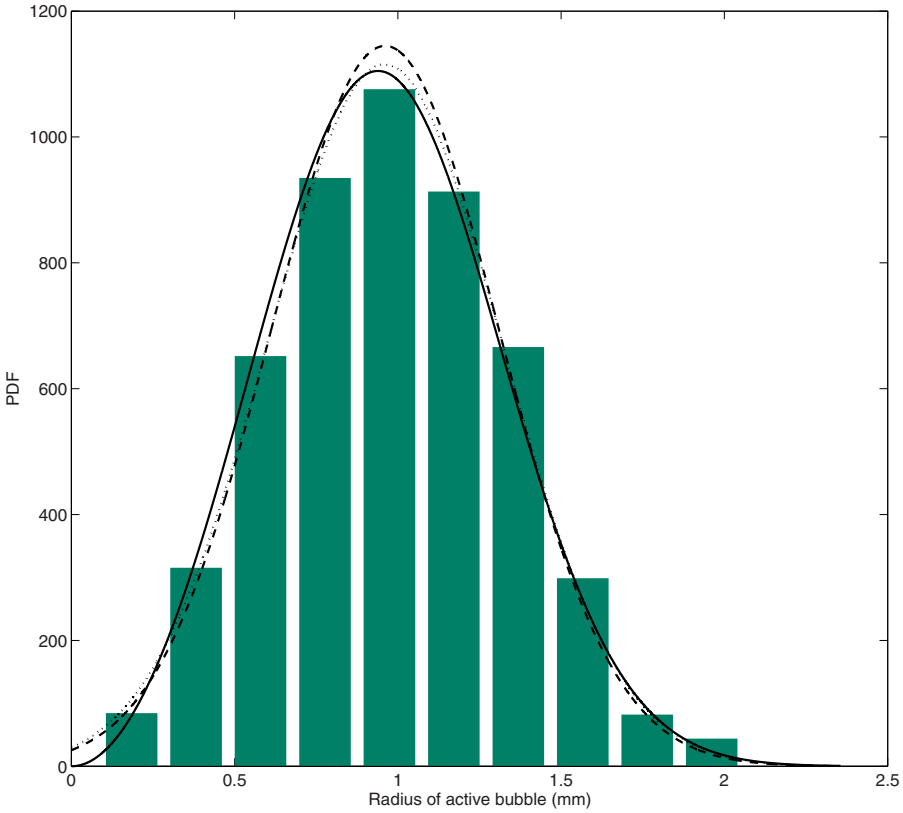


FIG. 1. (Color online) Histogram of active bubble sizes from a Monte Carlo computer simulation based on about 250 000 randomly drawn active growth stages and random final-stage sizes according to the measured PDFs. The solid line shows the theoretical result, Eq. (17), the dashed line, for comparison, a normal distribution with same mean and standard deviation, and the dotted line a normal distribution based on the measured sizes of the final-stage bubbles.

tributed final-stage sizes, all the moments can be computed in terms of the mean $\mu_f=\overline{a_f}$ and the standard deviation $\sigma_f=(\overline{a_f^2}-\overline{a}_f^2)^{1/2}$ only. In this case, we find

$$\overline{a} = \frac{3}{4}\frac{\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4}{\mu^3 + 3\mu\sigma^2}$$

$$\overline{a^2} = \frac{3}{5}\frac{\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4}{\mu^3 + 3\mu\sigma^2}. \tag{18}$$

The final-stage mean and the standard deviation can be inverted from Fig. 4 of Wahlberg and Westerberg[6] using linear free-field bubble theory to yield $\mu_f=0.96$ mm and $\sigma_f=0.36$ mm. Because of the small sample size, it cannot firmly be decided if the distribution is normal. However, the hypothesis that it is normal cannot be rejected at the 5% confidence level. Based on this data, the final-stage bubble size distribution does not seem to be dependent on the pressure, i.e., depth. Assuming that the sizes of the released bubble mainly depend on the physiology of the herring, this is the expected result. Using these numbers, we find $\overline{a}=0.96$ mm and a standard deviation of $\sigma=0.35$ mm for the full distribution of all active bubbles. At first, this seems paradoxical but, in fact, the upper half of the final-stage distribution, as hinted above, just compensates for the growth stages. The almost precise match of the numbers is a coincidence owing to the particularly measured final-stage sizes. Figure 1 shows the distribution of active bubble sizes based on the experimentally available data on final-stage radii.

The second case is simpler, since no final-stage PDF needs to be considered. The analysis then leads to $\overline{a}=(3/4)a_f$ and $\overline{a^2}=(3/5)a_f^2$.

The mean active bubble size can be used in evaluating Eq. (16) to approximate the full bubble size average since the deviations are only by about 10%. However, we use the actual distribution in the following analysis. Assuming normally distributed acoustically active bubbles, the bubble size averaged spectrum, Eq. (15), has characteristic attributes. It peaks slightly below the resonance frequency of the mean active bubble sizes, falls off sharply toward lower frequencies, and more gently toward higher frequencies, as is shown in Fig. 2. The particular production mechanism proposed, in which $\langle \dot{V}^2 \rangle$ is held constant as opposed to, for example, the relative initial distortions, is responsible for a less peaked spectrum and the more gradual falloff toward higher frequency, which follows a −2.8 power law.

Wilson *et al.*, in the electronic appendix to their publication, provide a sample herring chirp. Although not sufficient to truly test the ensemble averaged power spectral density discussed above, this nevertheless can be used to provide a consistency check of the overall picture. Figure 3 shows the shape of the spectrum of this single pulse and gives for comparison our −2.8 power-law curve for the high frequency tail, which overall compares favorably. This is an encouraging result given the fact that it was obtained on Atlantic Herring in a different environment, attesting to the general applicability of our model. Field data of noise from herring schools containing many pulses give a more convincing statistical picture and is presented in the subsequent paper. We also note that the first peak occurs where it is predicted to occur. It is not hard to imagine that if sufficiently many individual pulses are summed with slightly varying active

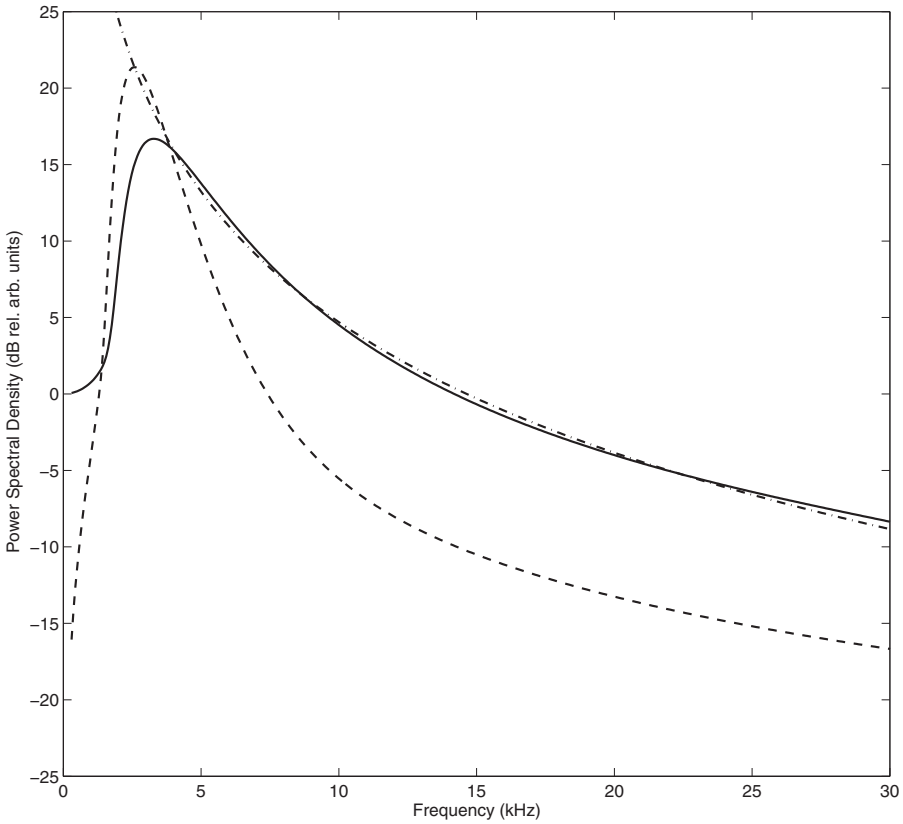


FIG. 2. Power spectral density according to the individual fish chirp model, Eq. (15), and normally distributed active bubble sizes with $\bar{a}=0.96$ mm and $\sigma=0.35$ mm (solid line). A fitted power-law curve that matches the high-frequency tail of the power spectral density is indicated with the dash-dotted line. The dashed line shows for comparison the same spectrum, however, with fixed $\langle(a_e/a_0)^2\rangle$.

bubble sizes, a smooth power spectral density, such as shown in Fig. 2, will emerge.

III. RESULTS AND DISCUSSION

The model estimates of the power spectral density from Eq. (16) can directly be used to interpret both the tank *and* field observations made by Wahlberg and Westerberg. Let us first consider their LPC experiments: Of the four factors of Eq. (16), two, the SL and the bubble dynamics term, are known from the data. The pulse rate $\dot{n}$ is known in order of magnitude, and the effective flow rate is unknown. Wahlberg and Westerberg reported an average duration of 4.2 s for a

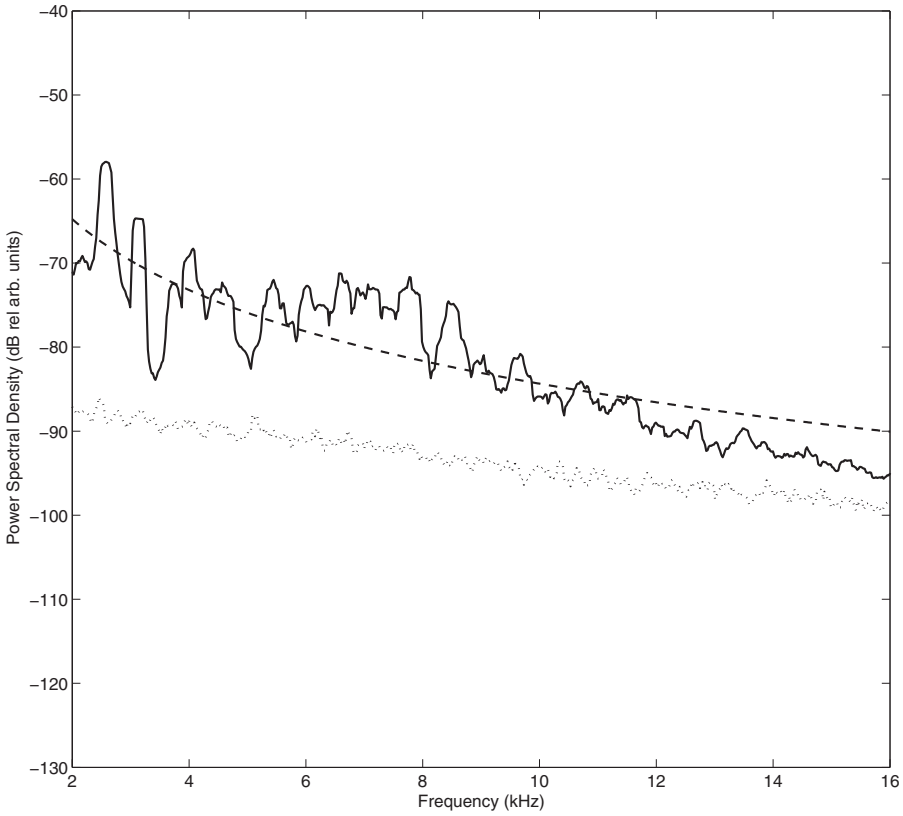


FIG. 3. Measured power spectral density of a *single* herring chirp (solid line), recorded by Wilson *et al.*, compared to the predicted *ensemble* averaged spectral power law (dashed line) resulting from the acoustic model of individual herring gas release. The dotted line indicates the estimated background of the measurements reported by these authors.

pulsed chirp corresponding to the gas release of an individual fish. These chirps are found to contain "up to hundreds" of pulses, each of them emitted due to a single excitation of the observed bubble. The authors do not report the precise pulse rate for the LPC case, but the field data indicated $\dot{n} \simeq 200\ \mathrm{s}^{-1}$. Hence, let us assume that for the tank experiment $\dot{n} \approx 50-200\ \mathrm{s}^{-1}$. The SL of 73 ± 8 dB with regard to μPa rms reference 1 m, measured by Wahlberg and Westerberg, can be used in Eq. (16) to estimate the effective rms flow rate $\langle \dot{V}^2 \rangle^{1/2}$. Using our assessment of active bubble radii discussed in the previous section for normally distributed final-stage sizes, the bubble dynamics term of Eq. (16) evaluates to about $4.6 \times 10^5\ \mathrm{s}^{-1}$ for the environmental conditions in Wahlberg and Westerberg's LPC. Inverting from the measured SL for the effective flow rate into the bubble, we find $\langle \dot{V}^2 \rangle^{1/2} \simeq 0.9$ ml/min for the tank conditions. The estimated standard deviation errors are +2.3 ml/min and −0.6 ml/min leading to a predicted lower bound of 0.3 ml/min and an upper bound of 3.2 ml/min for the effective flow rate. These account for the measured error in the sound levels, for the uncertainties in the pulse rate, as well as for potential discrepancies between the true flow rate and the effective flow rate of our model, where we assume an uncertainty of 30%. According to this estimate, the average total volume of gas released during the pulse is smaller than one-quarter of a milliliter. Given this amount, O(10) bubbles of the observed average final-stage sizes can be filled and released during the period of the chirp. The vast majority of the excitations must occur while the bubble remains attached to the ventral side of the herring, which yields the observed frequency decrease in the spectra. In extreme cases, with large final-stage bubble sizes, the number of bubbles can be O(1) or O(100) in the other extreme.

The average length of their herring specimen was measured to be 19.8 cm. According to Nero[21] *et al.*, herring of this size have an equivalent swim bladder radius of about 1 cm at the sea surface, equivalent to a bladder volume of about 4 ml. When exposed to pressures of 0.2 bar of their LPC (at which the gas release was observed to be initiated), following Boyle's law, the swim bladder mean radius increases to 1.63 cm. This corresponds, as noted by Wahlberg and Westerberg, to an increase in bladder volume by a factor of about 5. For the fish to remain neutrally buoyant, four times the original bladder volume, amounting to about 17 ml, would have to be ejected, if the bladder was initially filled. Direct physiological measurements of bladder volume of captive herring by Blaxter and Batty,[22] however, indicate much smaller volumes of about 0.75 ml, possibly indicating that, under tank conditions, the bladder is not fully inflated. (In their natural habitat, this is accomplished by gulping of air at the sea surface.) A bladder that is initially only partially full could explain why, in the LPC experiments of Wahlberg and Westerberg, the static pressure had to be decreased to 0.2 bar to trigger gas release. At this pressure, the partially filled bladder had expanded from the volume measured by Baxter and Batty to the assumed full volume of about 4 ml, as observed by Nero *et al.* Any further volume increase could consequently have triggered the venting of bladder gas.

Using the predictions of our model, we find a total released gas volume of up to about 0.22 ml. This is consistent with the above reasoning and it is noteworthy that a volume of the correct order of magnitude resulted purely from an inversion of the acoustic emissions of the herring without any reference to bladder size. Clearly, there are also other possible explanations for the prediction of this rather low volume; for example, the possibility that gas released by the specimen happens in a series of many pulses over a longer period of time. Wahlberg and Westerberg did not focus on the total amount of released gas. Their description of the experimental procedures does not seem to indicate this possibility. (The observation time might, however, not have been sufficiently long.)

As a side note, we want to point out that Eq. (14) allows an assessment of the magnitude of the bubble oscillations from the effective flow rate. Even in the case of strongest observed intensities, we find that $a_\varepsilon / a_0 \leqslant 0.001$, validating our small oscillation approach.

Let us now turn to the interpretation of the *field data* recorded by Wahlberg and Westerberg. In this case, the authors understandably do not report the average SL. However, they give a synopsis of the phenomenological signal features and a single example of a chirp. From this, out of the four unknown quantities of our model, all but the SL are known or can be assessed. The average pulse rate, as has been mentioned, is $\dot{n} \simeq 200\ \mathrm{s}^{-1}$. The average effective flow rate can be estimated from the pulse length and the initial and final bubble sizes to be $\langle \dot{V}^2 \rangle^{1/2} \simeq 2.5$ ml/min. Consistent with their featured sample pulse, which we reproduce in Fig. 4, this assumes that the total number of bubbles released is of O(1). This assumption is also supported by the fact that the pulse duration for the field data signal is much smaller than for the LPC setup. According to this assumption, we take the simplified bubble size average that applies for single bubble release, which to some degree is a convention. If the average were performed over signals containing the noise of many gas-releasing fish, the relevant size distribution would still be Eq. (17). In any case, the difference between the two schemes in term of SL (dB) is negligible. Putting all this together, we predict an average SL for individual herring in the field of about 89 ± 10 dB with regard to μPa rms reference 1 m, the error considering the uncertainties in the measurements of Wahlberg and Westerberg as well as the possibility that in some cases two bubbles are released.

The sample chirp displayed by Wahlberg and Westerberg and reproduced here in Fig. 4 can also be analyzed in the same fashion as their average field data. The total number of pings, from visual inspection of the spectrogram, is about 27 over the 100 ms period. The final and initial bubble sizes estimated from the frequency decay are about 0.7 mm and 1.5 mm, respectively. The final-stage frequency from their spectrogram seems to be a bit lower but has been chosen here in accordance with the reported higher limits. This yields an average effective flow rate of 9 ml/min and a SL of about 100 dB with regard to μPa rms reference 1 m, which is on the upper side of the expected levels in the field.

Our model predicts a significantly higher SL in the field as compared to the LPC. This hints that the behavioral basis

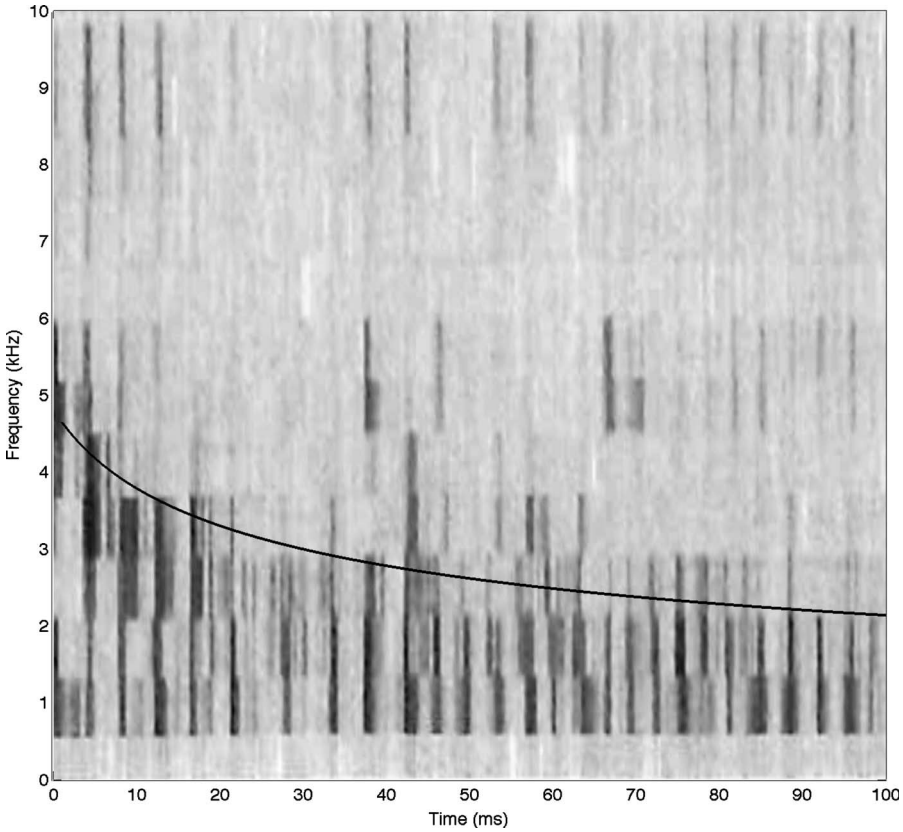


FIG. 4. Spectrogram of a single selected sample pulse recorded at a spawning aggregation of herring (reproduced from Wahlberg and Westerberg, 2003) overlaid with the expected center frequency decay that is based on a the effective flow rate estimate from the initial and final bubble size.

of the gas release was quite different between the two situations. In the LPC, the herring was forced to release gas to adapt to new buoyancy requirements, simulating a vertical migration behavior. The field data taken, as described by Wahlberg and Westerberg, were on spawning herring in shallow water, late at night. Based on the known migratory behavior of herring, the school can reasonably be assumed to have adapted to the local pressure environment and the gas release was likely be triggered by agitation due to predators or mating behavior.

Given that our theoretical model predictions for the gas release noise production by herring corroborates and complements the observations made by other researchers, our next step will be to estimate the noise levels produced by adult herring *schools*. Initially, we will focus on the noise production during their diel vertical migration and during predator avoidance. If these simulations suggest that a passive detection of these sounds is feasible, we will make field observations to verify our predictions and initiate experiments to use passive acoustics to monitor the herring in winter embayments.

Finally, the limited amount of data suggests that herring is capable of controlling the flux of released gas over almost one order of magnitude under similar environmental conditions. The spread of the observed SL under constant and controlled conditions as well as observations of acoustic emissions of herring at constant depths[23] further supports this hypothesis. This strongly suggests that gas release, as observed in natural herring, serves not solely the mechanistic adjustment of bladder volume to static buoyancy requirements but is also available to the species for more sophisticated behavioral modes, such as communication and more active modes of predator escape and avoidance.[2,6,22,24] This is of particular interest to us and relates directly to observations of intense nighttime feeding by marine mammals on the adult herring in the winter.[3] One of the primary predators of adult herring in the winter is the humpback whale. Humpback whales are known for the elaborate songs,[25] for their group hunting techniques, their use of bubble nets to herd herring into a tight school, and possibly for their use of acoustics for acoustic detection[26] and tracking of the herring schools. The use of air bubbles to herd and capture herring has also been visually observed in killer whales (*Orcinus Orca*) during cooperative feeding activities close to the surface.[27–29] Given that the herring appear to often have excess gas in their bladders and potentially the ability to finely control its release, we hypothesize that bubbles are actively used to avoid "acoustic predators." This hypothesis seems plausible, noting that the gas release not only reduces the target strength (TS) of the herring, making it less detectible to acoustic predators, but also has the potential to masks their route of escape, since released bubble clouds could potentially have a larger TS than the escaping fish.

ACKNOWLEDGMENTS

The authors would like to thank Magnus Wahlberg for kindly sharing a sample of his data, Dick Thorne for many helpful discussions and financial support through a grant from the NMFS, the associate editor K. G. Foote for carefully managing the peer review, and Otis Brown, the dean of the Rosenstiel School, for his assistance and support in de-

veloping this multidisciplinary line of research. This work was supported by Dr. Ellen Livingston, the Office of Naval Research.

[1]G. L. Thomas and R. E. Thorne, "Acoustical-optical assessment of pacific herring and their predator assemblage in Prince William Sound, Alaska," Aquat. Living Resour. **16**, 247–253 (2003).
[2]R. E. Thorne and G. L. Thomas, "Acoustic observations of gas bubble release by pacific herring (Clupea-Harengus-Pallasi)," Can. J. Fish. Aquat. Sci. **47**, 1920–1928 (1990).
[3]G. L. Thomas and R. E. Thorne, "Night-time predation by Steller sea lions," Nature (London) **411**, 1013–1013 (2001).
[4]L. Nottestad and B. E. Axelsen, "Herring schooling manoeuvres in response to killer whale attacks," Can. J. Zool. **77**, 1540–1546 (1999).
[5]M. A. Bishop and S. P. Green, "Predation on Pacific herring (Clupea pallasi) spawn by birds in Prince William Sound, Alaska," Fish Oceanogr. **10**, 149–158 (2001).
[6]M. Wahlberg and H. Westerberg, "Sounds produced by herring (Clupea harengus) bubble release," Aquat. Living Resour. **16**, 271–275 (2003).
[7]B. Wilson, R. S. Batty, and L. M. Dill, "Pacific and Atlantic herring produce burst pulse sounds," Proc. R. Soc. London, Ser. B **271**, S95–S97 (2004).
[8]T. G. Leighton, P. R. White, C. L. Morfey, J. W. L. Clarke, G. J. Heald, H. A. Dumbrell, and K. R. Holland, "The effect of reverberation on the damping of bubbles," J. Acoust. Soc. Am. **112**, 1366–1376 (2002).
[9]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[10]M. J. Buckingham, *Noise in Electronic Devices and Systems* (Halsted, Chichester, 1983), p. 372.
[11]M. S. Longuet-Higgins, "An analytical model of sound production by raindrops," J. Fluid Mech. **214**, 395–410 (1990).
[12]M. S. Longuet-Higgins, "Monopole emission of sound by asymmetric bubble oscillations. 1. Normal-modes," J. Fluid Mech. **201**, 525–541 (1989).
[13]M. S. Longuet-Higgins, "Monopole emission of sound by asymmetric bubble oscillations. 2. An initial-value problem," J. Fluid Mech. **201**, 543–565 (1989).
[14]M. Nicholas, R. A. Roy, L. A. Crum, H. Oguz, and A. Prosperetti, "Sound emissions by a laboratory bubble cloud," J. Acoust. Soc. Am. **95**, 3171–3182 (1994).
[15]M. S. Longuet-Higgins, B. R. Kerman, and K. Lunde, "The release of air bubbles from an underwater nozzle," J. Fluid Mech. **230**, 365–390 (1991).
[16]G. B. Deane and M. D. Stokes, "Scale dependence of bubble creation mechanisms in breaking waves," Nature (London) **418**, 839–844 (2002).
[17]G. B. Deane and M. D. Stokes, "The acoustic signature of bubbles fragmenting in sheared flow," J. Acoust. Soc. Am. **120**, EL84–EL89 (2006).
[18]M. S. Longuet-Higgins, "The crushing of air cavities in a liquid," Proc. R. Soc. London, Ser. A **439**, 611–626 (1992).
[19]E. Lamarre and W. K. Melville, "Void-fraction measurements and sound-speed fields in bubble plumes generated by breaking waves," J. Acoust. Soc. Am. **95**, 1317–1328 (1994).
[20]F. Minnaert, "On musical air bubbles and the sounds of running water," Philos. Mag. **16**, 235–248 (1933).
[21]R. W. Nero, C. H. Thompson, and J. M. Jech, "In situ acoustic estimates of the swimbladder volume of Atlantic herring (clupea harengus)," ICES J. Mar. Sci. **61**, 323–337 (2004).
[22]J. Blaxter and R. Batty, "The herring swimbladder: Loss and gain of gas," J. Mar. Biol. Assoc. U.K. **64**, 441–459 (1984).
[23]A. L. Schwarz and G. L. Greer, "Responses of pacific herring, clupea-harengus-pallasi, to some underwater sounds," Can. J. Fish. Aquat. Sci. **41**, 1183–1192 (1984).
[24]J. Blaxter and J. Hunter, "The biology of clupeids," Adv. Mar. Biol. **20**, 1–233 (1982).
[25]W. W. L. Au, A. Frankel, D. A. Helweg, and D. H. Cato, "Against the humpback whale sonar hypothesis," IEEE J. Ocean. Eng. **26**, 295–300 (2001).
[26]L. N. Frazer and E. Mercado, "A sonar model for humpback whale song," IEEE J. Ocean. Eng. **25**, 160–182 (2000).
[27]L. Nottestad, A. Ferno, and B. E. Axelsen, "Digging in the deep: killer whales' advanced hunting tactic," Polar Biol. **25**, 939–941 (2002).
[28]L. Nottestad, A. Ferno, S. Mackinson, T. Pitcher, and O. A. Misund, "How whales influence herring school dynamics in a cold-front area of the Norwegian Sea," ICES J. Mar. Sci. **59**, 393–400 (2002).
[29]T. Simila and F. Ugarte, "Surface and underwater observations of cooperatively feeding killer whales in northern Norway," Can. J. Zool. **71**, 1494–1499 (1993).

Correlation analysis of three-dimensional strain imaging using ultrasound two-dimensional array transducers

Min Rao[a] and Tomy Varghese
Department of Medical Physics, The University of Wisconsin-Madison, 1300 University Avenue, 1530 MSC, Madison, Wisconsin 53706

(Received 24 January 2008; revised 4 June 2008; accepted 5 June 2008)

Two-dimensional (2D) transducer arrays represent a promising solution for implementing real time three-dimensional (3D) ultrasound elastography. 2D arrays enable electronic steering and focusing of ultrasound beams throughout a 3D volume along with improved slice thickness performance when compared to one-dimensional (1D) transducer arrays. Therefore, signal decorrelation caused by tissue motion in the elevational (out-of-plane) direction needs to be considered. In this paper, a closed form expression is derived for the correlation coefficient between pre- and postdeformation ultrasonic radio frequency signals. Signal decorrelation due to 3D motion of scatterers within the ultrasonic beam has been considered. Computer simulations are performed to corroborate the theoretical results. Strain images of a spherical inclusion phantom generated using 1D and 2D array transducers are obtained using a frequency domain simulation model. Quantitative image quality parameters, such as the signal-to-noise and contrast-to-noise ratios obtained using 1D, 2D, and 3D motion tracking algorithms, are compared to evaluate the performance with the 3D strain imaging system. The effect of the aperture size for 2D arrays and other factors that affect signal decorrelation are also discussed. [DOI: 10.1121/1.2953310]

PACS number(s): 43.80.Qf [FD] Pages: 1858–1865

I. INTRODUCTION

Elastography is an imaging modality that is based on mapping local internal strains that tissues experience after a quasistatic or dynamic deformation.[1–3] Until recently, most studies devoted to elastography were focused on the estimation of the axial component of the strain. However, motion in tissue occurs in three dimensions and is coupled with each other when subjected to an axial compression or deformation. Although the third component of the displacement vector can be obtained from two-dimensional (2D) motion estimation by using the assumption of incompressibility, 2D motion tracking may suffer from decorrelation noise caused by out-of-plane tissue movement in the elevational direction. In addition, the incompressibility assumption may not hold in some tissues,[1,2] leading to errors. In addition, the Poisson's ratio may not be constant, such as for poroelastic tissue when edema is present.[3,4] Therefore, all the components of the strain tensor and displacement vector are required to characterize the resulting deformation.[5,6] Because the components of the strain tensor are coupled, accurate estimations of all components are necessary for a complete visualization of the strain incurred in tissue. Without those components, other important parameters such as shear strains cannot be estimated. In addition, knowledge of all the strain tensor components leads to an accurate Young's modulus reconstruction of the underlying tissue elasticity.[7,8]

To estimate displacement vectors and strain tensors in all three dimensions, three-dimensional (3D) volume data are required. 2D transducer arrays represent a promising solution for implementing real time 3D data acquisition. 2D arrays provide the capability for electronic steering and focusing of ultrasound beams throughout a 3D volume. Thus, they have improved slice thickness attributes when compared to one-dimensional (1D) and multirow transducer arrays. A problem facing the development of 2D arrays is the complexity arising from the large number of array elements required in such transducers and the channel capacity needed for the ultrasound system. Because of the complexity of implementing 2D array systems, it is particularly important to develop theoretical models to evaluate the performance and tradeoffs of strain imaging obtained with 2D arrays and to determine optimum parameters for these systems.

Previous works on 3D strain imaging have been reported in the literature.[9–13] Promising 3D strain imaging results have been demonstrated using 1D arrays with mechanical translation in the elevational direction utilized to acquire multiple B-scans followed by reconstruction to create the 3D image.[10,12] Konofagou and Ophir[9] described a 3D motion tracking algorithm and presented simulated strain images obtained using 1.5D and 2D array transducers. In that work, however, the simulated pre- and postdeformation ultrasonic radio frequency (rf) signals were obtained through a convolution of the point-spread function (PSF) with a tissue scattering function, where the 3D beam forming procedure with a 2D array was not simulated. Awad and Yen[11] also described a 3D strain imaging system using a prototype rectilinear sparse 2D array. However, thus far the performance of 3D strain imaging systems using fully sampled 2D arrays has not been thoroughly investigated.

Signal decorrelation noise is one of the major limiting factors in strain estimation and imaging.[8,9] The normalized

[a]Electronic mail: minrao@wisc.edu

correlation coefficient has been utilized previously to ascertain the degree of signal decorrelation and to quantify the accuracy and precision of strain estimates. A large amount of theoretical derivations has been reported in the literature to calculate the correlation coefficient between the pre- and postdeformation rf signals.[4,5,10–12] In most of these publications, 2D models were utilized since ultrasound beams with 1D array transducers are significantly wider in the elevational direction when compared to the lateral direction. Thus, scatterer movement in the elevational direction did not cause a significant loss of coherence in the echoes. In the case of 2D arrays, however, the out-of-plane motion (in the elevational direction) needs to be considered since the beam width is similar in both the lateral and elevational directions. Therefore, a 3D model is required to evaluate signal decorrelation due to the 3D motion. Kallel and Ophir[5] proposed a 3D model for predicting the correlation coefficient using separable point-spread functions in the axial, lateral, and elevational directions. In that work, the expression for the correlation coefficient was derived under the assumption that the lateral and elevational components of displacement are constant within the ultrasound beam. In practice, tissue scatterers move relative to each other in 3D space due to the applied compression or deformation. Thus the resulting signal decorrelation is caused by tissue distortion in all three dimensions. Therefore, the assumption of a simple translation of tissue scatterers in the lateral and elevational directions during compression is not valid and may result in a biased estimation of the correlation coefficient.

In this paper, a closed form theoretical expression is derived for the correlation coefficient between pre- and postdeformation ultrasonic rf signals. Signal decorrelation due to the 3D motion of scatterers within the ultrasonic beam has been considered. To corroborate the theoretical development, ultrasound (US) simulations using 1D and 2D arrays are performed and the results of a simulated uniformly elastic phantom, which verify the theoretical expression, are presented. Factors that affect signal decorrelation are discussed. Another purpose of this study is to present strain images from a simulated spherical inclusion phantom obtained using 1D and 2D arrays with different aperture sizes. Modern 3D beam forming techniques with 2D array transducers, such as dynamic aperture and dynamic receive focusing, have been implemented in our simulation.

II. THEORY

Ultrasonic rf echo signals before and after the applied deformation is modeled using

$$s_1 = P(x,y,z) \otimes T(x,y,z),$$

$$s_2 = P(x,y,z) \otimes T(x/b, y/b, z/a), \tag{1}$$

where subscripts 1 and 2 denote pre- and postdeformation echo signals obtained from an elastic tissue medium, and x, y, and z are the lateral, elevational, and axial coordinates, respectively, $T(x,y,z)$ denotes the tissue scattering function, and $P(x,y,z)$ is the pulse-echo PSF of the imaging system. The symbol $\otimes$ denotes the convolution operation. The parameters a and b denote the compression or expansion factor that scales the tissue scattering function defined in terms of the actual applied tissue strain ε

$$a = 1 - \varepsilon, \quad b = 1 + \frac{\varepsilon}{\nu}, \tag{2}$$

where ν is the Poisson ratio. There are several models that define the $T(x,y,z)$ term. To simplify the tissue model, we assume a large number of very small inhomogeneities (Rayleigh scatterers) with respect to the wavelength of the PSF,[14]

$$T(x,y,z) = \sum_i T_i \delta(x - x_i, y - y_i, z - z_i), \tag{3}$$

where $\delta(x,y,z)$ is the 3D Dirac or impulse function, (x_i, y_i, z_i) denote the randomly distributed centers of each inhomogeneity, and T_i represents the echogenicity of each scatterer. The scatterer distribution is assumed to be δ correlated, which means the correlation length of scatterers is very short compared to the acoustic wavelength of the transmit pulse.

The cross correlation between the signals acquired before and after deformation can be written as[15]

$$\langle s_1 s_2^* \rangle = \int\int\int P(l_{1x}, l_{1y}, l_{1z}) P(l_{2x}, l_{2y}, l_{2z}) \exp(j\Delta\phi) dx dy dz, \tag{4}$$

where $\Delta\phi$ is the phase difference between pre- and postdeformation signals, which can be written as $\Delta\phi = 4\pi(R_2 - R_1)/\lambda$, where λ is the wavelength at the center frequency, and R_1 and R_2 are the distances from the scatterer to the transducer for the pre- and postdeformation situations, respectively,

$$R_1 = \sqrt{(x - x_0)^2 + (y - y_0)^2 + z^2},$$

$$R_2 = \sqrt{(bx - x_0)^2 + (by - y_0)^2 + (az)^2}. \tag{5}$$

Under the assumption that the depth of interest z_{01} is much larger than the beam width, we obtain

$$R_2 - R_1 \approx \frac{(b^2-1)x^2 - 2(b-1)x_0 x + (b^2-1)y^2 - 2(b-1)y_0 y + (a^2-1)z^2}{2z_{01}}. \tag{6}$$

Here we assume that the 3D PSF is separable in three dimensions, an assumption that holds at the focus,[16] We can write $P(x,y,z)$ as

$$P(x,y,z) = p_x(x)p_y(y)p_z(z). \tag{7}$$

The axial, lateral, and elevational PSF components may be modeled as Gaussian envelopes with characteristic widths σ_z, σ_x, and σ_y, respectively,[17]

$$p_z(z) = \exp(-z^2/2\sigma_z^2), \quad p_x(x) = \exp(-x^2/2\sigma_x^2),$$

$$p_y(y) = \exp(-y^2/2\sigma_y^2). \tag{8}$$

If we assume that the incident pulses are Gaussian shaped with a center frequency f_c and a 6 dB fractional bandwidth B, the characteristic width of the axial PSF component can be computed using

$$\sigma_z = \frac{c}{2\pi B f_c} = \frac{\lambda}{2\pi B}. \tag{9}$$

For 2D arrays, the beam widths along the lateral and elevational directions are determined by the effective transducer aperture D, the wavelength, and the depth, which can be approximated by

$$\sigma_x = \sigma_y = \frac{0.4 z_{01}\lambda}{D}. \tag{10}$$

The lateral, elevational, and axial distances from scatterers to the beam axis for the pre- and postdeformation cases can be written as

$$l_{1x} = x - x_0, \quad l_{1y} = y - y_0, \quad l_{1z} = z - z_{01},$$

$$l_{2x} = bx - x_0, \quad l_{2y} = by - y_0 \quad l_{2z} = az - z_{02}, \tag{11}$$

where z_{02} is the depth of the point of interest after deformation. We can then evaluate the integral in Eq. (4) along the axial, lateral, and elevational directions separately,

$$\langle s_1 s_2^* \rangle = I_x I_y I_z, \tag{12}$$

where I_x, I_y, and I_z are the absolute values of the integral along x, y, and z directions,

$$I_x = \left| \int \exp\left(-\frac{l_{1x}^2 + l_{2x}^2}{2\sigma_x^2}\right) \times \exp\left(j\frac{2\pi[(b^2-1)x^2 - 2(b-1)x_0 x]}{\lambda z_{01}}\right) dx \right|$$

$$= \sqrt{\frac{2\pi}{b^2+1}}\sigma_x \exp\left[-\frac{(b-1)^2 x_0^2}{2\sigma_x^2(b^2+1)}\right]. \tag{13}$$

Similarly we can obtain

$$I_y = \sqrt{\frac{2\pi}{b^2+1}}\sigma_y \exp\left[-\frac{(b-1)^2 y_0^2}{2\sigma_y^2(b^2+1)}\right], \tag{14}$$

$$I_z = \sqrt{\frac{2\pi}{a^2+1}}\sigma_z \exp\left[-\frac{(az_{01}-z_{02})^2}{2\sigma_z^2(a^2+1)} - \frac{(a^2-1)^2}{2B^2(a^2+1)}\right]. \tag{15}$$

The correlation coefficient can be computed as follows:

$$\rho = \frac{\langle s_1 s_2^* \rangle}{\sqrt{\langle s_1 s_1^* \rangle \langle s_2 s_2^* \rangle}}. \tag{16}$$

Substituting Eq. (12) into Eq. (16), we obtain

$$\rho = \rho_x \rho_y \rho_z, \tag{17}$$

where ρ_x, ρ_y, and ρ_z are the effective correlation coefficients along the lateral, elevational, and axial directions, respectively,

$$\rho_x = \sqrt{\frac{2b}{b^2+1}} \exp\left[-\frac{(b-1)^2 x_0^2}{2\sigma_x^2(b^2+1)}\right], \tag{18}$$

$$\rho_y = \sqrt{\frac{2b}{b^2+1}} \exp\left[-\frac{(b-1)^2 y_0^2}{2\sigma_y^2(b^2+1)}\right], \tag{19}$$

$$\rho_z = \sqrt{\frac{2a}{a^2+1}} \exp\left[-\frac{(az_{01}-z_{02})^2}{2\sigma_z^2(a^2+1)} - \frac{(a^2-1)^2}{2B^2(a^2+1)}\right]. \tag{20}$$

Equation (17) provides an expression for the correlation coefficient between the pre- and postdeformation signals acquired from the location of interest. For elastographic processing, finite gated segments of the echo signal are utilized. A gated data segment is generally selected using a rectangular window. Generally, the same window is used for both the pre- and postdeformation data segments. Attenuation and focusing effects can be included in the window function $w(t)$, such that the intensity of $s(t)$ can be considered to be constant with depth. Thus, the correlation coefficient between the pre- and postdeformation data segments can be computed using[18]

$$\rho_{1,2} = \frac{\int_{t_1}^{t_2} w^2(t)\langle s_1(t)s_2^*(t)\rangle/\bar{I}dt}{\int_{t_1}^{t_2} w^2(t)dt} = \frac{\int_{-Z/2}^{Z/2} w^2(\xi)\langle s_1(\xi)s_2^*(\xi)\rangle/\bar{I}d\xi}{\int_{-Z/2}^{Z/2} w^2(\xi)d\xi} = \frac{\int_{-Z/2}^{Z/2} w^2(\xi)\rho(\xi)d\xi}{\int_{-Z/2}^{Z/2} w^2(\xi)d\xi}, \tag{21}$$

where $\bar{I} = \langle ss^* \rangle$ is the mean signal intensity, Z is the window length of the windowed rf echo signal segment, and ξ is the distance from a position within the window to the center of the window, which is in the range from $-Z/2$ to $Z/2$. $\rho(\xi)$ is the correlation of signals resulting from position ξ within the window of the rf segment. To compute $\rho(\xi)$, we only need to change the value of z_{01} and z_{02} as

$$z_{01} = z_0 + \xi, \quad z_{02} = az_0 + \xi, \tag{22}$$

where z_0 is the depth of the rf segment center. Substituting Eq. (22) into Eq. (20), and then plugging the result into Eq. (17), we can obtain $\rho(\xi)$.

III. SIMULATION

A. Method

A uniformly elastic simulated TM phantom and a single inclusion phantom with dimensions $4\times4\times4$ cm^3 was utilized to evaluate the signal decorrelation due to deformation. Both phantoms were constructed using the FEA software (ANSYS Inc., Canonsburg, USA), with a Poisson's ratio of 0.495 assumed for both the inclusion and the background. The inclusion phantom contains a 1 cm diameter spherical inclusion, which was three times stiffer than the background. To apply a uniaxial deformation, the simulated phantoms were fixed on the top surface and a displacement load (in the z direction) was applied to the bottom surface. Tissue displacement in the lateral and elevational direction were unconstrained. The local displacement due to the applied deformation was then estimated by solving the associated partial differential equations numerically.

The FEA displacement field was then utilized in an ultrasound simulation program[19,20] to generate pre- and postdeformation rf echo signal data for elastographic processing. This program simulates the frequency domain response of ultrasound wave propagating through a scattering medium. The frequency response is then transformed back to the time domain to obtain ultrasound radio frequency frames. This simulation program achieves similar rf waveforms when compared to typical time domain simulation programs such as FIELD II.[21] Modern beam forming techniques such as apodization, dynamic aperture, dynamic receive focusing, and 3D beam steering can also be simulated with our simulation program.[20] The accuracy of the frequency domain simulation model has also been assessed in our previous work[20] by comparing it with FIELD II. In our simulation, a numerical phantom was constructed with scatterers modeled using 50 μm radius glass beads, which were randomly distributed in the phantom at a number density of 10 scatterers/mm^3 to ensure Rayleigh scattering.[16] The 3D displacement fields from ANSYS, after appropriate interpolation to a regular Cartesian grid, were used to displace the scatterers and generate the deformed phantom to generate the postdeformation signals.

A 1D linear array was modeled, which consisted of 64 elements with dimensions of 150 μm$\times$12 mm and center-to-center spacing of 0.2 mm. This configuration produces a lateral aperture size around 12.8 mm. The elevational focus was set at 50 mm and the lateral transmit focus at 30 mm. Dynamic receive focusing was also utilized with an F-number set at 2. Simulated 2D array transducers having 32×32 and 64×64 active elements were also modeled with square elements of size 0.15 mm with a 0.2 mm center-to-center element separation. Hence, the aperture size is around 6.4 and 12.8 mm for 32×32 and 64×64 2D arrays, respectively. We used a single transmit focus at 30 mm and dynamic receive focus with an F-number set at 2 for both the lateral and elevational directions. The incident pulses were modeled to be Gaussian shaped with a 6 MHz center frequency and a 60% bandwidth. The simulations were performed assuming that the sound speed in the phantom was constant at 1540 m/s and negligible attenuation. The sampling frequency utilized was 52 MHz. For a $4\times4\times4$ cm^3 simulated phantom, 200 elevational slices were generated. In each elevational slice, 400 A-lines were simulated in the lateral direction. So a total of 80 000 (400×200) A-lines were acquired in one single 3D scan.

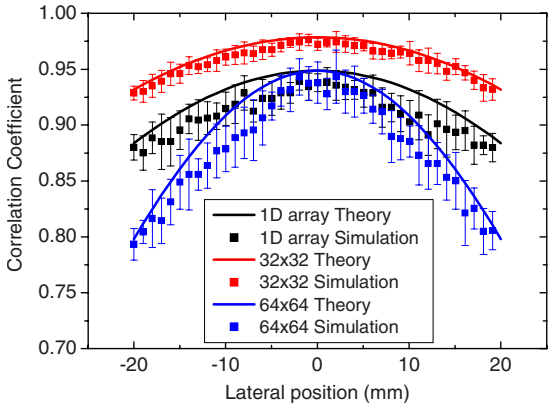


FIG. 1. (Color online) Comparison between theoretical prediction and simulation results for the correlation coefficient of pre- and postcompression rf segments as a function of the lateral position. Results are shown for (a) 1D array and (b) 32×32 and (c) 64×64 2D transducer arrays. The applied strain was 2%.

B. Results

Theoretical results demonstrating the variation in the correlation coefficient are verified using simulated rf data obtained using a uniformly elastic phantom. Figure 1 shows a comparison between the theoretical prediction and simulation results for the correlation coefficient of pre- and postdeformation rf segments as a function of lateral position. Results obtained using the 1D array and 32×32 and 64×64 2D array transducers are presented. The correlation coefficient was obtained using a 3 mm rf data segment centered at a depth of 3 cm from the transducer and at an elevational distance of 1 cm from the axis of symmetry of the elevational displacement field. The applied strain was 2%. The simulation results for the normalized correlation coefficient are measured using cross-correlation techniques applied to data segment from pre- and postdeformation rf A-lines. The error bars denote the standard deviation of the mean correlation coefficient estimates over 12 independent data sets, which were obtained using 12 independent realizations of the ultrasound simulation program with randomly distributed tissue scatterers. For the 1D array, we assume that the beam profile has a Gaussian shape in the elevational direction as well. To predict the correlation coefficient, the beam width used in Eq. (8) was obtained by measuring the simulated ultrasound beam profiles.

As illustrated in the figure, the correlation coefficient decreases with an increase in the lateral position moving from the center to the edge of the phantom. This is due to the increased scatterer motion or displacement across the beam at the edges of the phantom when compared to the motion or displacement at the center. Note that the correlation coefficient decreases faster for the 64×64 2D array when compared to the 1D array and 32×32 array. This difference is due to the different aperture sizes utilized for the three simu-

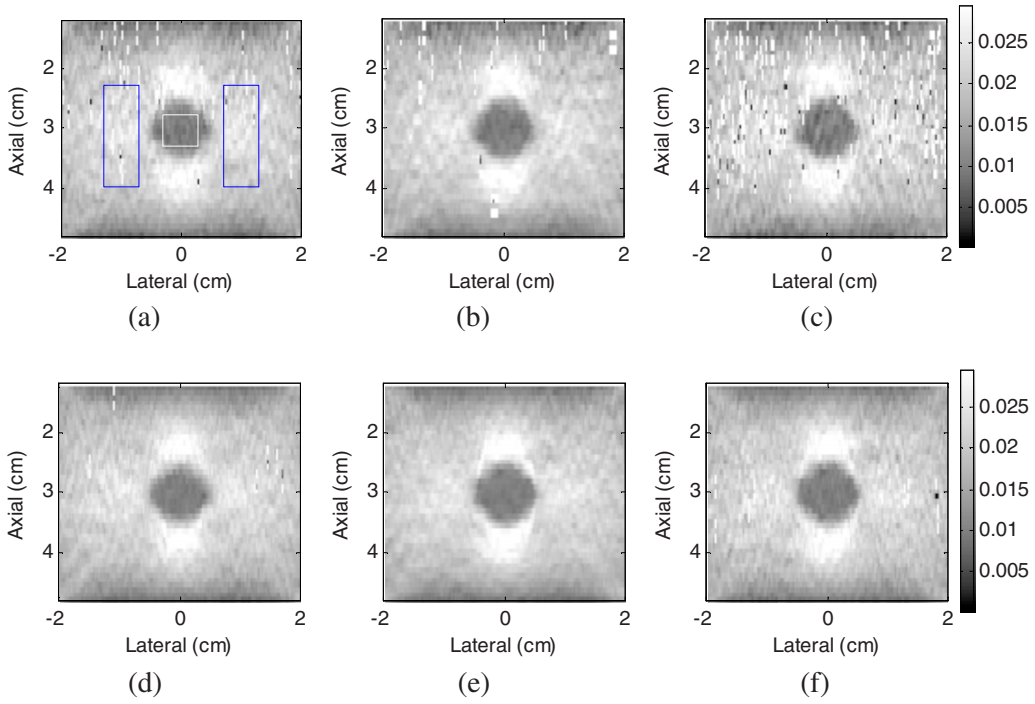


FIG. 2. (Color online) Simulated axial strain images of a spherical inclusion phantom obtained using [(a) and (d)] 1D array [(b) and (e)] 32×32 and [(c) and (f)] 64×64 2D arrays. The first row presents strain images obtained using 2D motion tracking algorithm without motion compensation in the elevational direction, and the second row depicts strain images obtained using 3D motion tracking algorithm. The center frequency is 4 MHz.

lated cases. Larger aperture sizes produce narrower beams and thereby induce faster signal decorrelation.

Furthermore, it can be observed in Fig. 1 that the theoretical predictions of the correlation coefficient are slightly higher than the simulation results. This is because our theoretical model assumes that the displacement along the axial direction has been accurately tracked, in which case the correlation coefficient achieves the maximum value at that position. The simulation results, however, inevitably contain errors in the displacement estimates due to the motion tracking algorithm and sampling precision. Therefore, the correlation coefficient obtained from simulation is generally lower than that from theoretical prediction.

Figure 2 presents simulated strain images of the spherical inclusion phantom obtained using the 1D array [(a) and (d)] and 32×32 [(b) and (e)] and 64×64 [(c) and (f)] 2D arrays. The first row illustrates strain images obtained using 2D motion tracking algorithm without motion compensation in the elevational direction, and the second row depicts strain images obtained using a 3D motion tracking algorithm. In this study, we used a 1D kernel, which is approximately 3 mm in the axial direction and one *A*-line along the lateral direction, to compute the cross-correlation function and to determine the displacement. A three-point, least squares strain estimator[22] was then used to generate local strain estimates. To better visualize the decorrelation noise in the strain images, median filters were not used for the displacement and strain estimated. It is assumed that the spherical inclusion is centered at a depth of 3 cm from the transducer and a distance of 1.5 cm from the elevational axis (y_0=1.5 cm). The applied strain is 2%, and the center frequency is 4 MHz with a 60% bandwidth. As expected, the decorrelation noise has been reduced after elevational motion compensation. The strain images obtained using the 64×64 2D array are more sensitive to signal decorrelation due to elevational motion. This is because the ultrasound beam produced by the 64×64 array is narrower along both the lateral and elevational directions. The noise performance of the 1D array is similar to that of the 32×32 2D array, but the inclusion looks smaller in the strain images [Figs. 2(a) and 2(d)] due to the poor slice thickness performance. In this example, the 32×32 2D array is the best choice for strain estimation as it provides good performance in terms of the noise level and spatial resolution.

Figure 3 compares the SNR_e and CNR_e of the simulated axial strain images of the inclusion phantom obtained using the 1D array and 32×32 and 64×64 2D arrays shown in Fig. 2. The quantitative SNR_e and CNR_e parameters were calculated using a rectangular region of interest (ROI) within the inclusion and in the background region, as shown in Fig. 2(a). Note that the image quality is significantly improved using the 3D motion tracking algorithm, especially for the 64×64 2D array.

Figure 4 presents simulated axial strain images of the inclusion phantom obtained using 32×32 (left) and 64×64 (right) 2D arrays shown in the lateral-elevation plane at a depth of 3 cm. It is assumed that the spherical inclusion is centered at a depth of 3 cm from the transducer and on the elevational axis (y_0=0 cm). Observe the decreased decorrelation noise at the elevational edges of the phantom obtained using the 3D motion tracking algorithm, especially for the 64×64 2D array.

To illustrate the impact of the center frequency, simulated axial strain images of the inclusion phantom were obtained using center frequencies of 6, 8, and 10 MHz for a 32×32 2D array, as shown in Fig. 5. Results obtained by

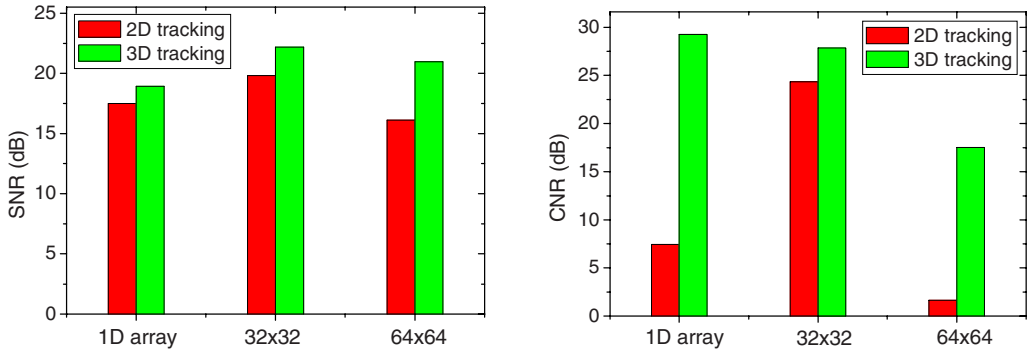


FIG. 3. (Color online) A bar graph of the signal-to-noise (SNR) and contrast-to-noise (CNR) values of the simulated axial strain images of the inclusion phantom obtained using 1D array and 32×32 and 64×64 2D arrays shown in Fig. 2.

applying a compression or deformation of 1% (top) and 2% (bottom) were compared using a 3D motion tracking algorithm. All the other simulation parameters were the same as those described in Fig. 4. As expected, increased decorrelation noise artifacts are observed when higher frequencies and larger deformations are used.

IV. DISCUSSION

2D array transducers represent a promising solution for the implementation of volume data acquisition for real time 3D strain imaging. The effective aperture size is an important parameter for 2D array systems since it determines the number of elements required for a given element size and spacing between adjacent elements. It is desirable to use a small number of active elements to obtain good strain images. Generally, larger apertures provide an improved ultrasound resolution but introduce increased decorrelation noise artifacts due to the narrower beam width. We need to take into consideration this tradeoff when choosing the aperture size of the 2D transducer array for strain imaging. Therefore, it is useful to understand how the aperture size and other system factors affect the correlation between pre- and postdeformation signals.

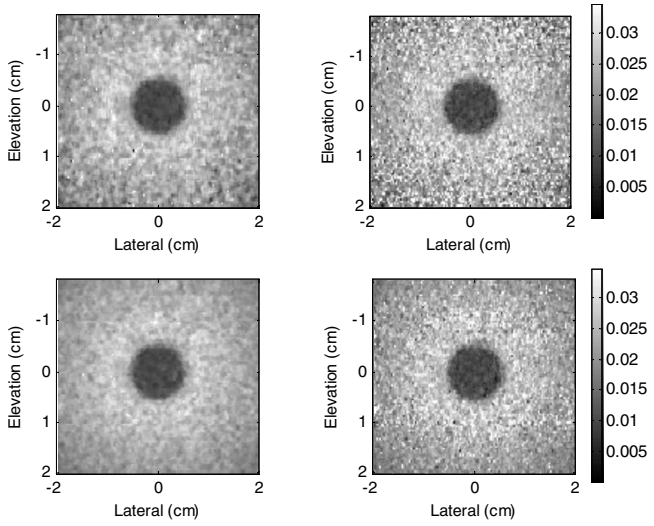


FIG. 4. Simulated axial strain images of the inclusion phantom obtained using 32×32 (left) and 64×64 (right) 2D arrays shown in the lateral-elevation plane at depth 3 cm. The first row presents strain images obtained using 2D motion tracking algorithm, and the second row depicts strain images obtained using 3D motion tracking algorithm. The center frequency is 4 MHz.

Figure 6 presents theoretical correlation coefficient curves as a function of the aperture size of a 2D array for different applied strains. The center frequency is 6 MHz and the fractional bandwidth is 60%. Results are obtained using 3 mm rf segments centered at a depth of 3 cm, in the center along the lateral direction and 1 cm from the center in the elevational direction (x_0=0 cm; y_0=1 cm). As illustrated in the figure, the decorrelation rate of the rf signal pairs increases with the value of the applied strain. This increased decorrelation rate is due to the fact that with increased strain more scatterers would leave the beam, leading to increased decorrelation, especially for larger aperture sizes.

Figure 7 shows the theoretical prediction of the correlation coefficient versus the 2D array aperture size for different center frequencies. We assume a fractional bandwidth of 60% for all the cases. Results are obtained using 3 mm rf segments centered at a depth of 3 cm, in the center along the lateral direction and 1 cm from the center in the elevational direction (x_0=0 cm; y_0=1 cm). The applied strain is 2%. As illustrated in the figure, the signal decorrelation rate of the rf signal pairs increases with the center frequency. This is primarily due to the fact that the ultrasound beam becomes narrower with the increased center frequency, enabling scatterers that are within the predeformation beam to leave the beam and newer scatterers to come within the beam after deformation when the beam becomes narrower, especially for larger aperture sizes.

Figure 8 shows the theoretical prediction of the correlation coefficient versus the 2D array aperture size for rf signals obtained at different elevational positions. Results are obtained using 3 mm rf segments centered at a depth of 3 cm, in the center along the lateral direction (x_0=0 cm). The applied strain is 2% and the center frequency is 6 MHz. As expected, the signal decorrelation rate is larger at the edges of the target, where the elevational motion is the largest. Signals obtained using larger aperture sizes are more sensitive to decorrelation caused by the elevational motion since the beam is narrower for the larger apertures. Note that the correlation coefficient is independent of the aperture size

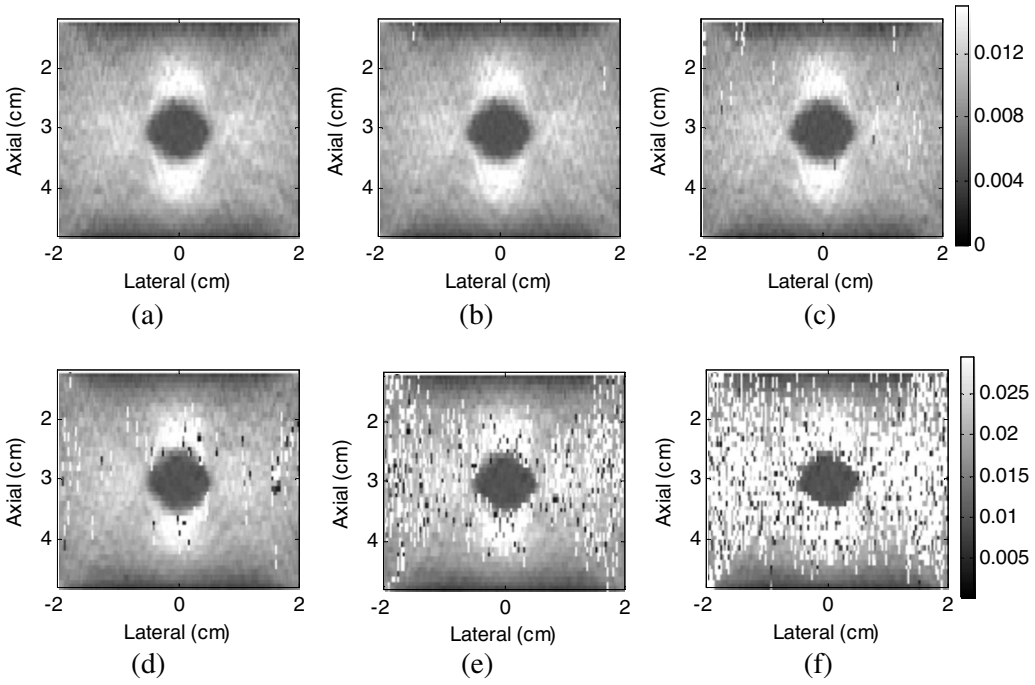


FIG. 5. Simulated axial strain images of the inclusion phantom obtained using the center frequency of [(a) and (d)] 6 MHz, [(b) and (e)] 8 MHz, and [(c) and (f)] 10 MHz for the 32×32 2D array. The applied strain is 1% (top) and 2% (bottom).

for rf signals obtained at the geometric center of the target, where the motion of tissue scatterers in the lateral and elevational directions is the smallest.

Figure 9 shows the theoretical prediction of the correlation coefficient versus the 2D array aperture size for rf signals obtained at different depths. Results are obtained using 3 mm rf segments in the center along the lateral direction and 1 cm from the center in the elevational direction (x_0 =0 cm; y_0=1 cm). The applied strain is 2% and the center frequency is 6 MHz. As illustrated in the figure, the signal decorrelation rate for the rf signal is higher at shallow depths. This is because at shallow depths the phase changes ($\Delta\phi$) are larger due to the motion of scatterers in the lateral and elevational directions. Note that our model estimates the correlation coefficient of rf segments after axial motion compensation. Thus, the correlation coefficient is higher regardless of the large axial displacement at deeper depths.

V. CONCLUSIONS

In this paper, we investigate the impact of decorrelation noise in strain imaging, resulting from the motion of tissue scatterers in all three dimensions. Under the assumption that the system PSF is separable, as shown in Eq. (7), a theoretical expression has been derived for the signal decorrelation between pre- and postdeformation rf echo signals. The theoretical prediction matches well with the simulation results. Simulated axial strain images of a spherical inclusion phantom obtained using both 1D and 2D array transducers are

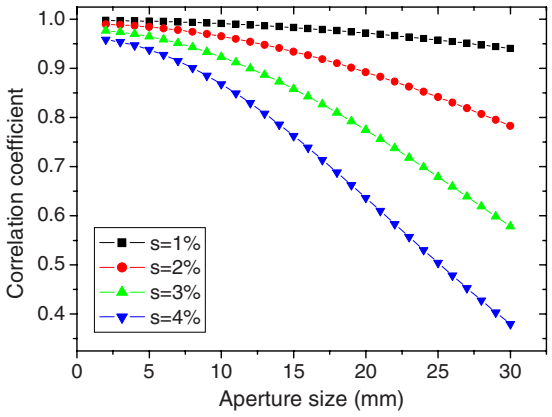


FIG. 6. (Color online) Plots of the theoretical prediction of the correlation coefficient vs the 2D array aperture size for different applied strains. Results are obtained using 3 mm rf segments centered at a depth of 3 cm, in the center along the lateral direction and 1 cm from the center in the elevational direction (x_0=0 cm; y_0=1 cm). The center frequency is 6 MHz.

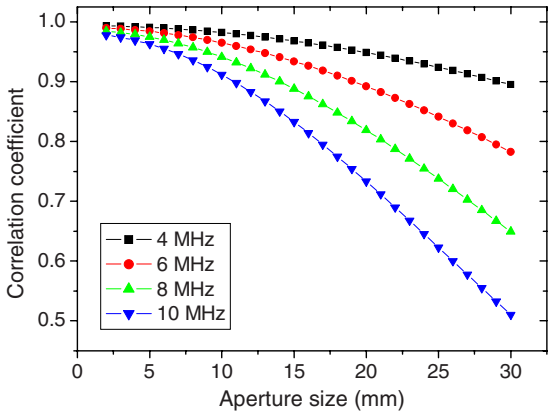


FIG. 7. (Color online) Plots of the theoretical prediction of the correlation coefficient vs the 2D array aperture size for different center frequencies. Results are obtained using 3 mm rf segments centered at a depth of 3 cm, in the center along the lateral direction and 1 cm from the center in the elevational direction (x_0=0 cm; y_0=1 cm). The applied strain is 2%.

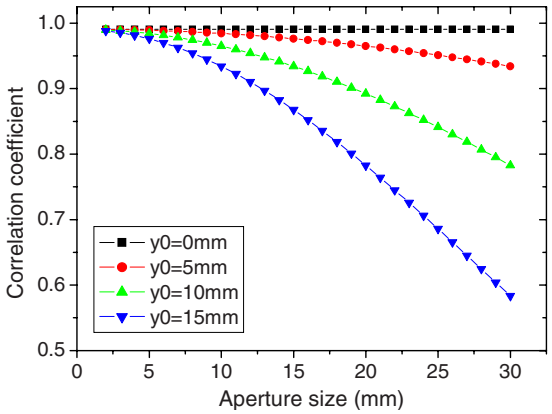


FIG. 8. (Color online) Plots of the theoretical prediction of the correlation coefficient vs the 2D array aperture size for rf signals obtained at different elevational positions. Results are obtained using 3 mm rf segments centered at a depth of 3 cm, in the center along the lateral direction (x_0=0 cm). The applied strain is 2% and the center frequency is 6 MHz.

also presented. A quantitative comparison of the image quality of strain images obtained using different motion tracking algorithms demonstrates that 3D motion tracking is necessary for imaging systems using 2D array transducers. For 3D real time strain imaging, the theoretical derivation presented in this paper would enable the estimation of optimal parameters, such as the aperture size, center frequency, and applied strain. To detect stiffer lesions at shallow depths, a 2D transducer array with a smaller effective aperture is a good choice since it provides high resolution and low decorrelation noise, and at the same time, fewer elements are required for a given spacing between elements.

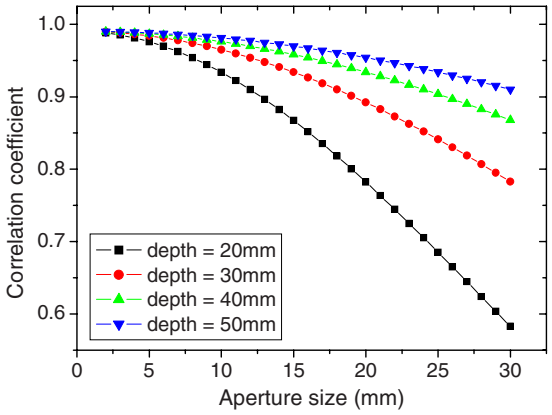


FIG. 9. (Color online) Plots of the theoretical prediction of the correlation coefficient vs the 2D array aperture size for rf signals obtained at different depths. Results are obtained using 3 mm rf segments in the center along the lateral direction and 1 cm from the center in the elevational direction (x_0 =0 cm; y_0=1 cm). The applied strain is 2% and the center frequency is 6 MHz.

[1]Y. G. Fung, *Biomechanical Properties of Living Tissues* (Springer, New York, 1981).
[2]J. S. Jurvelin, M. D. Buschmann, and E. B. B. Hunziker, "Mechanical anisotropy of the human knee articular cartilage in compression," Proc. Inst. Mech. Eng., Part H: J. Eng. Med. **217**, 215–219 (2003).
[3]E. E. Konofagou, T. P. Harrigan, J. Ophir, and T. A. Krouskop, "Poroelastography: Imaging the poroelastic properties of tissues," Ultrasound Med. Biol. **27**, 1387–1397 (2001).
[4]R. Righetti, J. Ophir, S. Srinivasan, and T. A. Krouskop, "The feasibility of using elastography for imaging the Poisson's ratio in porous media," Ultrasound Med. Biol. **30**(2), 215–228 (2004).
[5]F. Kallel and J. Ophir, "Three-dimensional tissue motion and its effect on image noise in elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1286–1296 (1997).
[6]M. Bilgen, "Dynamics of errors in 3D motion estimation and implications for strain-tenser imaging in acoustic elastography," Phys. Med. Biol. **45**, 1565–1578 (2000).
[7]M. Odonnell, A. R. Skovoroda, B. M. Shapo, and S. Y. Emelianov, "Internal displacement and strain imaging using ultrasonic speckle tracking," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 314–325 (1994).
[8]F. Kallel and M. Bertrand, "Tissue elasticity reconstruction using linear perturbation method," IEEE Trans. Med. Imaging **15**, 299–313 (1996).
[9]E. E. Konofagou and J. Ophir, "Precision estimation and imaging of normal and shear components of the 3D strain tensor in elastography," Phys. Med. Biol. **45**, 1553–1563 (2000).
[10]J. E. Lindop, M. T. Graham, A. H. Gee, and R. W. Prager, "3D elastography using freehand ultrasound," Ultrasound Med. Biol. **32**, 529–545 (2006).
[11]S. I. Awad and J. T. Yen, "3D strain imaging using a rectilinear 2D array," Ultrason. Imaging **29**, 220–230 (2007).
[12]A. V. Patil, C. D. Garson, and J. A. Hossack, "3D prostate elastography: algorithm, simulations and experiments," Phys. Med. Biol. **52**, 3643–3663 (2007).
[13]G. Said, O. Basset, J. M. Mari, C. Cachard, E. Brusseau, and D. Vray, "Experimental three dimensional strain estimation from ultrasonic sectorial data," Ultrasonics **44**, E189–E193 (2006).
[14]J. Meunier and M. Bertrand, "Ultrasonic texture motion analysis—Theory and simulation," IEEE Trans. Med. Imaging **14**, 293–300 (1995).
[15]R. F. Wagner, M. F. Insana, and S. W. Smith, "Fundamental correlation lengths of coherent speckle in medical ultrasonic images," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 34–44 (1988).
[16]R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," IEEE Trans. Sonics Ultrason. **30**, 156–163 (1983).
[17]M. Rao and T. Varghese, "Correlation analysis of the beam angle dependence for elastography," J. Acoust. Soc. Am. **119**, 4093–4101 (2006).
[18]Q. Chen, A. L. Gerig, U. Techavipoo, J. Zagzebski, and T. Varghese, "Correlation of RF signals during angular compounding," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 961–970 (2005).
[19]Y. D. Li and J. A. Zagzebski, "A frequency domain model for generating B-mode images with array transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 690–699 (1999).
[20]M. Rao, T. Varghese, and J. Zagzebski, "Simulation of ultrasound two-dimensional array transducers using a frequency domain model," Med. Phys. **35**, 3162–3169 (2008).
[21]J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267 (1992).
[22]F. Kallel and J. Ophir, "A least-squares strain estimator for elastography," Ultrason. Imaging **19**, 195–208 (1997).