CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 119, No. 2 February 2006

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Spherically focused capacitive-film, air-coupled ultrasonic transducer

Junho Song and D. E. Chimenti
Center for Nondestructive Evaluation, Iowa State University, Ames, Iowa 50011
and Department of Aerospace Engineering, Iowa State University, Ames, Iowa 50011
sjh4841@iastate.edu, chimenti@iastate.edu

Stephen D. Holland
Center for Nondestructive Evaluation, Iowa State University, Ames, Iowa 50011
sdh4@iastate.edu

Abstract: A spherically focused (no mirrors) capacitive-film, air-coupled ultrasonic transducer, constructed using a spherically deformed backplate and metalized polymer film, has been designed, fabricated, and its performance characterized. A 1 cm diameter device has a center frequency of 805 kHz and a 6 dB bandwidth of 760 kHz. Comparisons of field strength in the focal zone with theoretical calculations for a spherically focused piston show that the device achieves diffraction-limited focusing. The nominal focal point of 25 mm lies within 0.01 mm of the calculated value for this device.

PACS numbers: 43.38.Bs, 43.38.Gy [IBV]
Date Received: September 5, 2005 **Date Accepted:** October 16, 2005

1. Introduction

Recently in nondestructive evaluation (NDE), attention has centered on noncontact ultrasonic inspection methods because they are practical and efficient when the test article under inspection cannot be brought into contact with water. Currently, most air-coupled ultrasonic inspection methods utilize either conventional piezoceramic transducers or capacitive film transducers. When a solid piezoceramic transducer is used to couple sound into air, the large acoustic impedance mismatch between the element and air renders broadband matching nearly impossible.[1,2] Capacitive film transducers skirt the impedance mismatch problem by using a thin polymer film of low areal density as the vibrating element.[3] Biased to 100 V dc or more, this type of transducer functions by detecting the vibrational motion induced in the film by incident sound waves and converting this motion to electrical current.[1] The capacitive film transducer has a relatively low fabrication cost, high sensitivity, and very wide bandwidth compared with conventional piezoceramic transducers. Previous attempts to develop natively focused transducers of this type have met with only limited success. Approaches employing external devices, such as acoustic mirrors,[4] cylindrical focusing,[5] or Fresnel zone plates[6] have all been attempted. Unlike these previous approaches, our spherically focused capacitive transducer utilizes a spherical radiating surface, and therefore needs no mirror, zone plate, or any similar external device to effect focusing. In this paper, we present the development, fabrication, and testing of a spherically focused capacitive film air-coupled transducer, utilizing a spherically deformed backplate and conformal metalized polymer film in the shape of a spherical radiator.

2. Transducer construction

Our 10 mm spherically focused capacitive film transducer is fabricated with a 25.4 mm geometric focal length and an active angular sensitivity of ±15° with respect to the normal axis. It is designed to excite a large range of plate wave modes when in normal incidence for low-density engineering materials, such as plastics, carbon or glass–fiber composites, and lightweight alloys. The radiating surface of a fully constructed spherically focused capacitive film transducer is shown in Fig. 1(a).

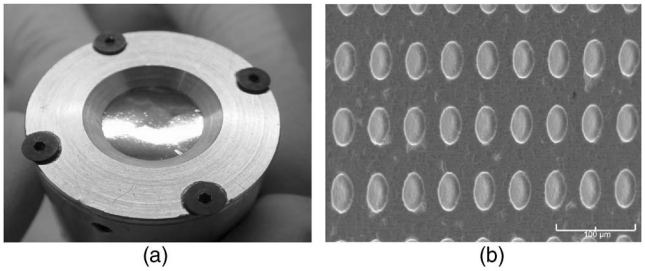


Fig. 1. (a) A photograph of a 5 mm radius, 25.4 mm focal length spherically focused capacitive micromachined air-coupled ultrasonic transducer. A white line on a 6 μm Mylar/Al film is the reflection of a light source. (b) A SEM image of a flexible copper/polyimide backplate.

Our concept for a spherically focused transducer is based on the premise that both sides of the capacitor, the backplate and the polymer film, can be made to conform to a section of spherically curved surface. We have accomplished this goal by utilizing a flexible copper(Cu)/polyimide(PI) backplate, molded to a spherically curved backplate fixture, and a suitably prepared conformal metalized Mylar film. The flexible and permanently deformable Cu/PI backplate is a two-layer structure consisting of a 17 μm thick copper layer bonded to a 130 μm thick polyimide substrate, commonly used for flexible printed circuits. We pattern the copper layer of this material with 40 μm depressions having 80 μm center-to-center spacing, as shown in Fig. 1(b). Then, the backplate is carefully deformed to conform to a machined spherically curved backplate fixture, whose radius is the same as the desired geometric focal length of the transducer.

A 6 μm thick aluminized Mylar film, mechanically deformed to give it a spherical shape, completes the transducer. The fabrication of a conformal polymer film was another innovation necessary for the construction of a natively focused capacitive air-coupled transducer. To suppress performance-robbing wrinkles in the Mylar film, we have mechanically stretched the metalized Mylar film using a warm steel ball bearing. The radius of the ball bearing is approximately the same as the geometric focal length of the spherically focused capacitive transducer. After stretching the Mylar film, it assumes a spherical shape and can be fitted directly to the Cu/PI backplate without wrinkling, when a bias voltage is also applied.

3. Transducer characterization

To characterize our new device, we measure the sound pressure fields radiating from the spherically focused transducer by a second 10 mm diameter capacitive film transducer, baffled by a 200 μm diameter aperture, giving a quasipoint receiver that is scanned through the focal zone of the spherical transmitter under study. The receiver uses the same film and construction details as the focused transmitter, so its bandwidth characteristics are identical to the probe under test. The receiver is biased to 200 V dc.

The focused probe is excited by a bandwidth-tailored 200 μs random-phase signal.[4] Figure 2(a) shows the typical response of our spherically focused capacitive transducer, and Fig. 2(b) shows its corresponding frequency spectrum. All our measurements are relative, so amplitude units are arbitrary. The latter shows that the frequency spectrum is centered at 805 kHz with a 6 dB bandwidth of approximately 760 kHz, which is measured at a lower and upper frequency of 446 and 1207 kHz, respectively. This bandwidth is not only far wider than all piezoceramic transducers, but also wider than most damped water-coupled piezoceramic transducers.

For reference we have defined a simple Cartesian coordinate system, as shown in Fig. 3(a). The origin of the coordinate system is located at the center of the concave face of the spherical backplate in the spherically focused capacitive transducer. Figure 3(a) shows the measured sound fields in the x-z plane at $y=0$, radiated from the spherically focused capacitive air-coupled transducer whose geometric focal length is 25.4 mm. The sound field is scanned in

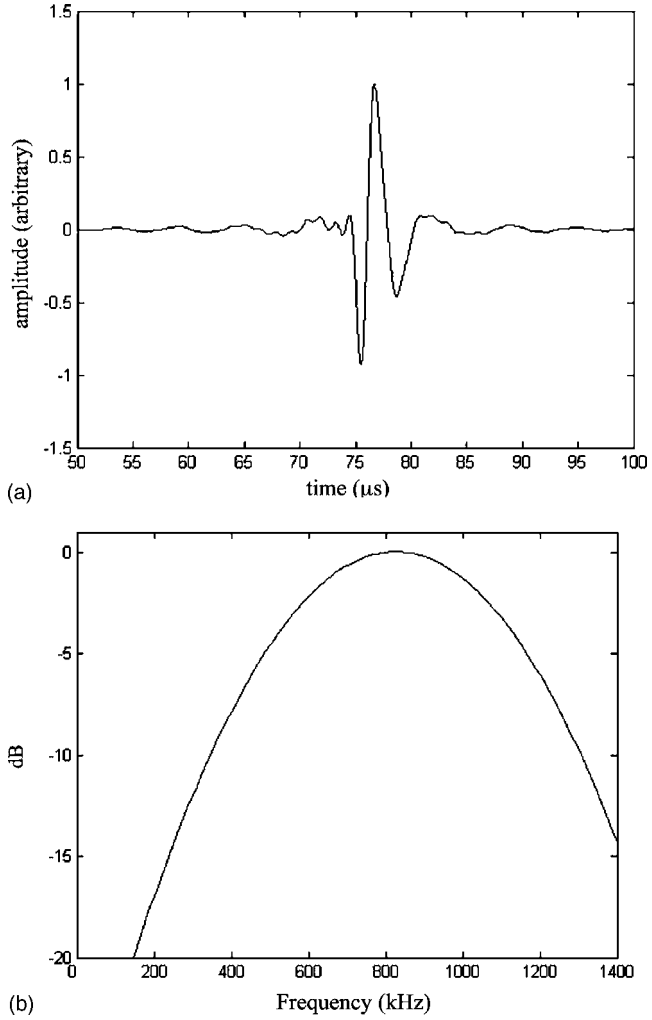


Fig. 2. (a) Typical amplitude response and (b) the corresponding frequency spectrum. The quasipoint receiver has a 200 V dc bias, and the transmitter is driven by a broadband excitation at 250 V peak to peak.

the *x*-*z* plane over an area of 8 mm × 35 mm and with spatial resolutions of 0.1 and 0.2 mm in the *x* and *z* axis, respectively. The figures show peak-to-peak sound field amplitudes at each point where dark red represents a much stronger sound field amplitude than a dark blue region. For broadband excitation, we have obtained the maximum amplitude at 24.9 mm. The focal zone extends from 17.1 and 34.1 mm, respectively. Figure 3(b) shows the measured sound field in the *x*-*y* plane at $z = 15$, 25, and 35 mm. The figure clearly shows a point focusing performance of the transducer.

To evaluate the transducer's performance, we compare our experimental result with a theoretical prediction using the Rayleigh–Sommerfield model.[7] Figure 4 shows the cross section of the focal region of the measured and theoretical sound pressure fields for a 10-cycle 800 kHz tone burst excitation, radiated from the spherically focused air-coupled transducer. The sound pressure from a focused piston radiator is

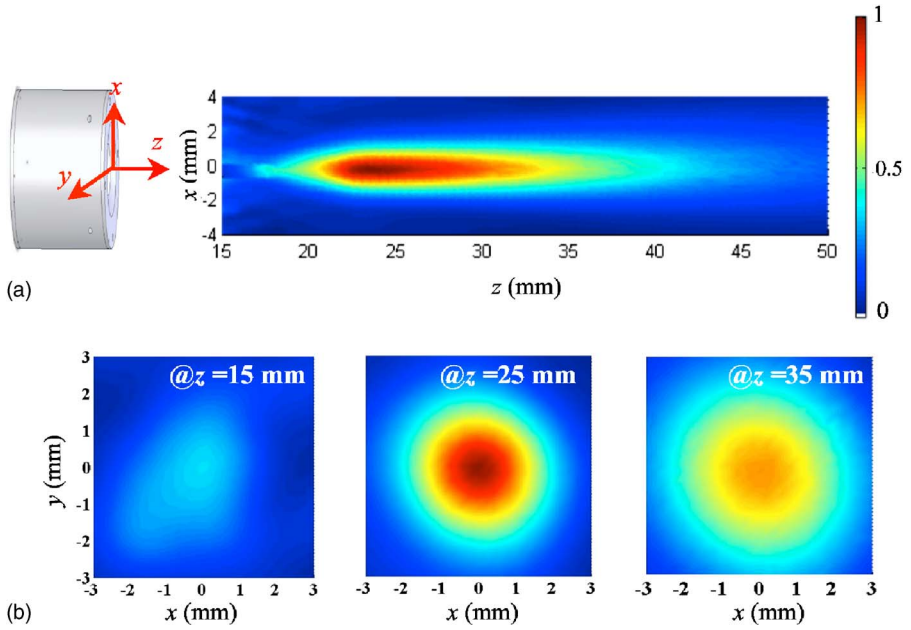


Fig. 3. (Color online) Measured sound pressure fields radiated from a 5 mm radius, 25.4 mm focal length spherically focused air-coupled transducer driven by broadband transient signals: (a) sound pressure field in the x-z plane at $y=0$, and (b) sound pressure fields in the x-y plane at $z=15, 25$, and 35 mm. Darker red regions represent stronger sound pressure fields than dark blue regions. (*Those looking at a printed black and white version of this figure may find it helpful to look at the color version of this figure in the online publication*).

$$p(R_0,y,\omega) = -i\omega\rho v_0 a^2[\exp(ik\bar{R}_0)/\bar{R}_0][J_1(kay/\bar{R}_0)/(kay/\bar{R}_0)], \quad (1)$$

where R_0 is the focal length, $\bar{R}_0=\sqrt{R_0^2+y^2}$, y is the radial distance, k is the wave number, ρ is the mass density of the medium, p is the radiating sound pressure, a is the radius of a piston

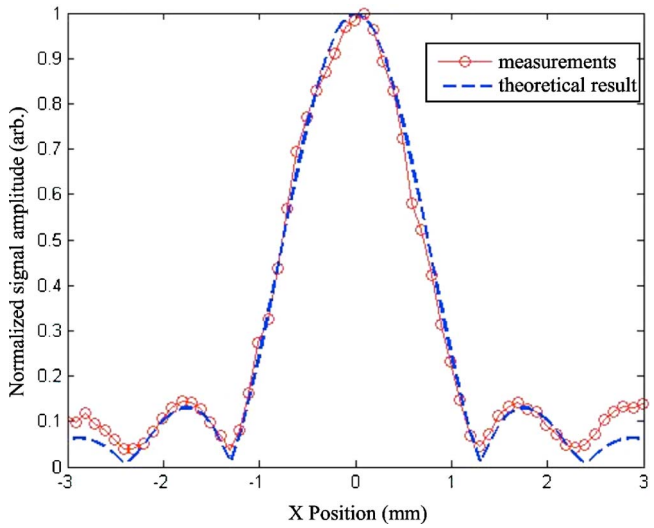


Fig. 4. (Color online) Cross sections of the focal region of the measured and theoretical sound pressure fields radiated from a 5 mm radius, 25.4 mm focal length, spherically focused air-coupled transducer when driven by an 800 kHz tone burst.

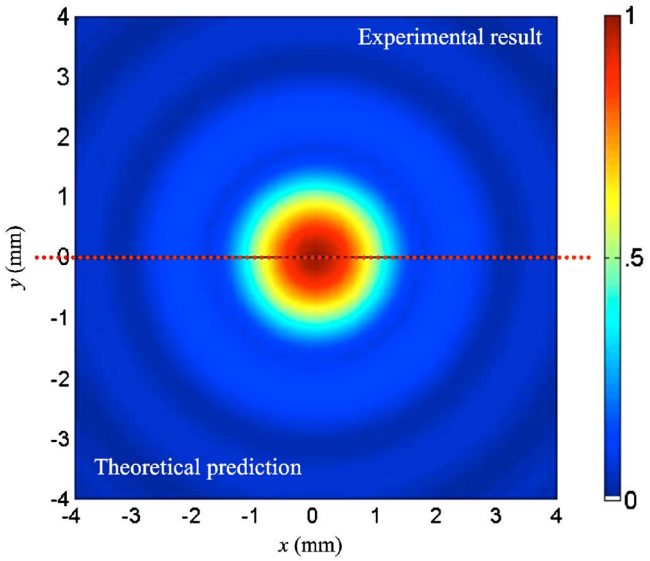


Fig. 5. (Color online) A comparison of the experimental measurement in the focal plane and Rayleigh–Sommerfeld theoretical prediction for a narrow band 500 kHz tone burst. The upper half is the experiment, and the lower half is the theoretical prediction, illustrating the transducer's nearly perfect axial symmetry. (*Those looking at a printed black and white version of this figure may find it helpful to look at the color version of this figure in the online publication*).

transducer, v_0 is the piston velocity (assumed uniform over the face of the radiator), and J_1 is the first-order Bessel function. The calculation has no adjustable parameters except for the arbitrary amplitude. Our measurements are obtained at the focal zone for each excitation signal, which we have found in the *x*-*z* plane scan. The full width at half-maximum (FWHM) value, or 6 dB dropoff point, is measured to be 1.38 mm, and its theoretical prediction is 1.37 mm. The theoretical prediction is sufficiently close to the experimental measurements for us to conclude that our device is operating like an ideal spherically focused piston radiator. Figure 5 shows the experimental measurement (in the upper half-frame) compared with the theoretical prediction (in the lower half-frame) of the focal-plane behavior of the focused transducer at 500 kHz. Because the radiating surface is spherical, resolution of the focused beam is diffraction limited, controlled only by deviations in the fixture sphericity, the device diameter, medium sound wave speed, and the frequency.

4. Conclusion

We have presented a simple, yet reliable, design of natively focused micromachined capacitive air-coupled ultrasonic transducers and have shown a simple method to fabricate them. By selecting a flexible substrate as a backplate, we eliminate one of the most difficult and unsolved problems in backplate fabrication. Moreover, because our device is natively focused, this transducer eliminates the need for auxiliary focusing devices, such as acoustic mirrors or zone plates. We have demonstrated that it behaves accurately like a spherically focused piston radiator. We anticipate this device's high signal amplitude, wide bandwidth, and optimal spatial resolution will significantly improve air-coupled ultrasonic nondestructive evaluation and imaging applications.

Acknowledgments

This work is supported by NASA under award NAG-1-029098.

References and links

[1]H. Carr and C. Wykes, "The performance of capacitive transducers," Ultrasonics **31**, 13–20 (1993).
[2]A. S. Ergun, A. Atalar, B. Temelkuran, and E. Özbay, "A sensitive detection method for capacitive ultrasonic transducer," Appl. Phys. Lett. **72**, 2957–2959 (1998).
[3]D. W. Schindel, D. A. Hutchins, L. Zou, and M. Sayer, "The design and characterization of micromachined air-coupled transducer capacitance," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 42–50 (1995).
[4]S. D. Holland, S. V. Teles, and D. E. Chimenti, "Air-coupled, focused ultrasonic dispersion spectrum reconstruction in plates," J. Acoust. Soc. Am. **115**, 2866–2872 (2004).
[5]T. J. Robertson, D. A. Hutchins, and D. R. Billson, "Capacitive air-coupled cylindrical transducers for ultrasonic imaging applications," Meas. Sci. Technol. **13**, 758–769 (2002).
[6]D. W. Schindel, A. G. Bashford, and D. A. Hutchins, "Focusing of ultrasonic waves in air using a micromachined Fresnel zone-plate," Ultrasonics **35**, 275–285 (1997).
[7]L. W. Schmerr, Jr., *Fundamentals of Ultrasonic Nondestructive Evaluation—A Modeling Approach* (Plenum, New York, NY, 1998), pp. 181–197.

Predicting perceived sharpness of broadband noise from multiple moments of the specific loudness distribution

Atsushi Marui[a)]
Multimedia Systems Laboratory, University of Aizu, Aizuwakamatsu-shi, Fukushima-ken 965-8580, Japan
atsushi.marui@mcgill.ca

William L. Martens
Schulich School of Music, McGill University, 555 Sherbrooke Street, Montreal, H3A 1E3, Canada
wlm@music.mcgill.ca

Abstract: Equal sharpness contours for broadband noise were generated through matches made between a standard reference noise and comparison noise stimuli varying in spectral envelope. Based upon the results of the sharpness matching task, a set of 20 percussive broadband noise stimuli was constructed and presented to three subjects in a sharpness rating experiment using white noise as a standard stimulus. Predicting obtained sharpness ratings for these percussive broadband noise stimuli from first and second moments of the stimulus specific loudness distribution was more successful than conventional prediction based only upon the weighted first moment.

PACS numbers: 43.66.Lj, 43.75.Gh, 43.75.Tv [DKW]
Date Received: July 22, 2005 **Date Accepted:** October 31, 2005

1. Introduction

Although the perceived sharpness of steady-state narrow-band noise may be predicted with good accuracy from the weighted first moment of the associated specific loudness distribution,[1] more complex broadband stimuli that vary in their overall spectral envelope are not as well predicted by the conventional model (see, for example, Ref. 2). In an effort to begin to bridge the gap between sharpness prediction for synthetic narrow-band noise stimuli and sound stimuli with more arbitrary spectra, the relationship between stimulus specific loudness distribution and perceived sharpness was assessed for percussive noise stimuli (with rapidly decaying amplitude envelopes) that differed in their overall spectral envelope. It has long been recognized that variation in overall spectral envelope is the main factor influencing sharpness;[3,4] however, given the simplicity of the sharpness prediction equation presented by Zwicker and Fastl,[1] this generally applicable equation does not provide a wholly adequate means for quantitative description of spectral envelope using parameters relevant to sharpness prediction. The two experiments reported in this paper were executed to provide psychoacoustic data that might aid in the formulation of a more successful prediction model for the perceived sharpness of broadband noise stimuli.

The conventional predictor of apparent sharpness described by Zwicker and Fastl,[1] which hereafter will be termed "Zwicker Sharpness," measures stimulus sharpness using a unit termed the *acum* (which means "sharp" in Latin). Zwicker Sharpness (ZS) of a given stimulus is referenced to the perceived sharpness of a narrow-band noise (one critical-band wide) with a center frequency of 1 kHz having a level of 60 dB sound pressure level (SPL). The perceived sharpness of this narrow-band stimulus establishes the 1 *acum* point on a perceptual scale for the sharpness of other stimuli. The numerical value S for the ZS is calculated as the weighted

[a)]Currently associated with Schulich School of Music, McGill University; corresponding author

first moment of the stimulus specific loudness distribution, a prediction of the auditory response distributed over critical bands. The equation is formulated as

$$S = 0.11 \frac{\int_0^{24\ \text{Bark}} N' g(z) z \, dz}{\int_0^{24\ \text{Bark}} N' \, dz} \text{acum}, \quad (1)$$

where N' is a specific loudness for each critical band z on a bark scale and $g(z)$ is a factor that places extra weight upon high frequencies (above 16 bark).

Though this equation was developed primarily to predict the perceived sharpness of narrow-band noise stimuli, it was also used by Zwicker and Fastl[1] to predict the relative sharpness of a broadband noise, which was found to be roughly twice as sharp as the narrow-band reference stimulus producing 1 *acum*. For their spectrally flat broadband noise stimulus, the calculated numerical value of ZS was nearly double that of the narrow-band noise stimulus, and this ratio was in close agreement with obtained psychoacoustic results. Since broadband noise stimuli were under investigation in the current study, a spectrally flat broadband noise was considered to serve as a better standard stimulus to be compared to other broadband noise stimuli differing in their overall spectral envelopes. Note that the stimuli that were employed in both experiments reported here contained significant energy at higher frequencies; therefore the results may not generalize beyond calculated ZS values ranging roughly between $S=1.6$ and $S=2.6$.

In contrast to previous studies of the perceived sharpness of noise stimuli in which band-widths were varied by manipulating cutoff frequencies,[1] the spectral envelopes of the stimuli used in the current study were varied using a pair of shelving filters that could independently boost the high- or low-frequency energy contained in each stimulus. The parameters of these shelving filters were based upon filters used to shape the tone color of audio inputs in modern music production and reproduction (described in detail in Ref. 2). For example, in practical terms all commercially available audio equipment, from guitar effect processors to home theater systems, have at least one control related to sharpness which does not alter the timbral features that allow different musical instruments to be distinguished. This control alters an attribute of timbre that related to the long-time-average spectrum, and is more typically associated with tone coloration. In this investigation of perceived sharpness variations in such audio processing, the focus is not on developing a comprehensive sharpness model to be used for predicting sharpness differences between different musical timbres (instruments). Rather, the investigation of more simple variations for a given instrumental timbre is under examination here. Although only the first moment of the critical-band distribution of specific loudness is used in calculating ZS, higher-order statistical moments of the specific loudness distribution might provide a more adequate characterization of the variation in spectral energy associated with the high- and low-frequency emphasis provided by the employed shelving filters. So the primary motivation for the current study was to determine whether any improvement in sharpness prediction could be observed by including in the prediction equation higher-order moments, such as spectral variance, skewness, and kurtosis. The first statistical moment (which is spectrally weighted in ZS calculation) is the first moment about the origin, or the mean of the critical-band distribution, also termed "spectral centroid." The second moment about the mean, here termed "spectral variance," measures dispersion centered at the spectral centroid. Other studies[5,6] have shown such spectral centroid and spectral variance to predict perceptual attributes of musical timbre using multidimensional scaling (MDS) on a number of musical instruments. Spectral skewness is defined as the third standardized moment, measuring asymmetry of the distribution; and spectral kurtosis is defined as the fourth standardized moment, measuring peakedness of the distribution with respect to the normal distribution. In the research

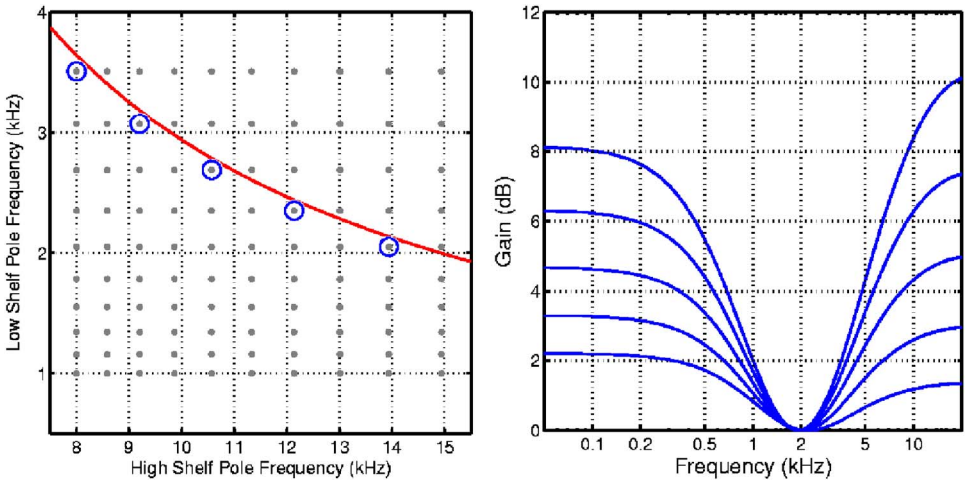


Fig. 1. Left: Small filled circles (dots) plot the filter pole coefficients used to produce sound stimuli that were employed in the sharpness matching task. The curve shows the log-linear regression line describing coefficients for the comparison stimuli found to match the standard stimulus in perceived sharpness. Open circles show coefficients for the five stimuli selected for the perceived sharpness rating experiment. Right: Filter gain for the five low-shelf and high-shelf filter settings associated with the selected five stimuli. Note that curves shown have been shifted to show 0 dB gain at 2 kHz for comparison, and differ from the relative spectral levels of the stimuli presented to the listeners, since prior to presentation, the overall gain of the filter for each stimulus was adjusted to give equal Zwicker Loudness.

described in this paper, the potential contributions of the four moments of the stimulus specific loudness distribution were evaluated using stepwise regression analysis, and the result was compared to predictions of perceived sharpness afforded by ZS.

2. Methods

2.1 Stimulus generation

Percussive noise stimuli were employed in both experiments reported here. They had temporal amplitude envelopes roughly matching a gunshot followed closely by short but dense reverberation. A single 600 ms synthesized noise sample was given a 1 ms raised-cosine rise and a 579 ms raised-cosine decay. Two first-order shelving filters were used to control, respectively, the relative amount of low-frequency and high-frequency content in a large set of otherwise identical stimuli. For the sharpness matching task described below, a 10×10 matrix of stimuli varying in tone coloration was generated by regularly sampling a two-dimensional control structure for the pole frequencies of the two shelving filters. For the shelving filters used in the subsequent sharpness rating task, the values chosen for their pole coefficients were based upon the results of the sharpness matching task.

The shelving filters used to synthesize the stimuli for the sharpness matching task were designed as follows. The low-frequency shelving filter had a zero fixed at 5 kHz and pole frequencies that ranged from 1–3.5 kHz in equal steps chosen to emulate critical-band spacing, along the ERB (equivalent rectangular bandwidth) scale. The zero frequency for the high-frequency shelving filter was fixed at 6 kHz, and its pole frequencies were uniformly spaced on the ERB scale ranging from 8–15 kHz. Ten examples of the spectral envelopes resulting from the combined filters are shown in Fig. 1 (right). The parameter that defined the low-frequency boost was the "low shelf pole frequency" that labels the ordinate of the graph shown in Fig. 1 (left). Here, lower frequency poles gave less low-frequency boost. In the case of the high-frequency shelving filter, lower frequency poles gave more high-frequency boost. Referring to Fig. 1 (right) will aid the reader in understanding the range of spectral envelope variation

achieved using these filters, the low-frequency boost reaching a gain extreme of around 8 dB at dc, and the largest high-frequency boost reaching an asymptotic gain of around 12 dB at the Nyquist frequency, 22.05 kHz.

Though perceived sharpness does not typically change much with changes in level,[1] the experimental stimuli were adjusted to give nearly the same loudness in order to simplify the listening tasks. This was not done through human loudness matching, but rather was accomplished automatically using an iterative procedure relying on a conventional instrumental measure for loudness, termed here "Zwicker Loudness" (described in detail in the Zwicker and Fastl text[1]). Zwicker Loudness is simply the sum of specific loudness over critical bands. The stimuli were digitally synthesized at an audio sampling rate of 44.1 kHz, and these audio signals were processed using a filter bank modeled after the human peripheral auditory system. The bank of filters was designed using modified version of Slaney's[7] Auditory Toolbox routine MAKEERBFILTERS. It may be questioned why an ERB-related program was used in calculating specific loudness, rather than bark. The reason for this choice was that the text by Zwicker and Fastl did not provide a clear statement regarding the shape of the analysis bands. However, the current model did employ their stated center frequencies and bandwidths for critical bands that were given in the text. The shapes that we assumed are those of the popular gammatone filter bank. The sound level in *sone* units was calculated using a formula found in Ref. 8: sones $=0.0631(I/I_0)^{0.03}$. For the current study, all stimuli were adjusted to give Zwicker Loudness values of 4 sone (corresponding to a loudness level of 60 phon). The stimuli were presented to the subjects over Sennheiser HD-590 headphones which were connected directly to the output of an SGI O_2. Informal evaluation by the experimenters confirmed that the experimental stimuli were nearly equal in perceived loudness.

2.2 Perceived sharpness matching

A two-alternative forced choice (2AFC) task was employed to track the low-frequency boost required to match the perceived sharpness of a variable comparison stimulus to that of a fixed-sharpness standard stimulus. The standard stimulus was a spectrally flat broadband noise presented in alternation with a comparison broadband noise stimulus that within a given block of trials had a constant high-frequency boost, but one of ten low-frequency boost levels. The pole frequencies for the two shelving filters that were factorially combined to generate the 10×10 matrix of synthesized stimuli are represented by the 100 small dots plotted on the graph shown in Fig. 1 (left). Within one block of trials tracking the point of subjective equality for sharpness of the standard and comparison stimuli, holding the high-frequency boost constant limited the variation to spectral envelopes associated with a single vertical column of dots plotted on this graph. Three male subjects experienced in such listening experiments completed five sharpness matches at each of the ten fixed levels of high-frequency boost in a single experimental session.

Because the values chosen for the pole coefficients of the shelving filters used to synthesize the stimuli for the subsequent sharpness rating task were based upon the results of the sharpness matching task, these results are briefly summarized here. The median values of the low-frequency boost required to produce a sharpness match were calculated for each of the ten fixed high-frequency boost settings, and for each of the three subjects. The best fitting log-linear regression line, calculated from the 30 median low-shelf pole frequency settings, had a slope of −0.9608 and an intercept of 7.3113. Since the scaling of the axes of the graph shown in Fig. 1 (left) is linear rather than log, the resulting log-linear regression line appears as a curve. The five open circles that appear in the figure mark the filter parameter values for five of 20 stimuli that were selected for the sharpness rating task. These five stimuli were chosen to be close to the log-linear regression line, and represent the median settings for stimuli that were chosen to have the same sharpness as the spectrally-flat standard stimulus. The zero frequencies for the five high-frequency shelving filters with fixed pole coefficients ranged from 8.0 to 13.9 kHz. The pole frequencies for the low-frequency shelving filters that were chosen to produce matching perceived sharpness ranged from 3.6 to 2.1 kHz for the five stimuli (and the associated spectral envelopes are shown in the right panel of Fig. 1). Three additional sets of stimuli were generated

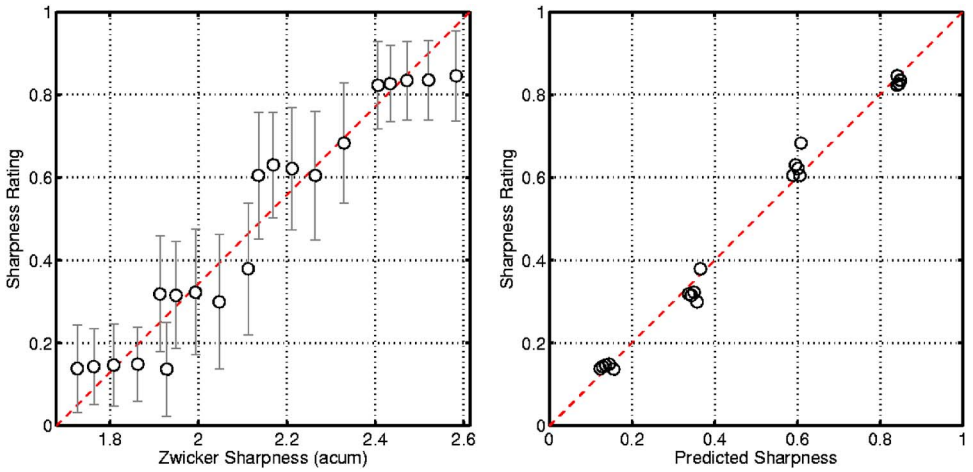


Fig. 2. Left: Open circles show mean sharpness ratings for the 20 stimuli plotted as a function of calculated ZS. Error bars show ranges of unity standard deviation. Right: The same mean sharpness ratings for the 20 stimuli, but plotted as a function of the values calculated using the two-term sharpness prediction function. No error bars were shown for the right figure to allow for a clearer view of the sharpness ratings. The dashed diagonal line in each plot shows the robust linear fit for each of the two sharpness prediction equations.

using filter parameters derived by multiplying these five by three constants, 0.44, 0.67, and 1.50. Thus two of the sets were generally darker than the five sharpness-matched stimuli, and one was generally brighter. This operation produced a set of 20 stimuli that exhibited a range of ZS values, extending roughly from $S=1.6$ to $S=2.6$. The 20 experimental sound stimuli, as well as the exact frequency values specifying them, can be found online.[9]

2.3 Perceived sharpness rating

Direct ratings of the perceived sharpness of the 20 experimental stimuli were made by the same three male listeners who had participated in the sharpness matching experiment. This experiment required the production of sharpness ratings on 15 sets of the 20 stimuli in a single session of 300 ratings, where 20 stimuli were ordered differently within each set. Listeners rated the perceived sharpness on a continuous scale using a slider on a graphical user interface (GUI) developed within the MATLAB programming environment. The GUI developed for this experiment allowed a listener to play each stimulus several times while adjusting the slider before the setting was recorded via a button press. Listeners completed a practice session of 30 ratings prior to the experimental session during which time they were instructed to focus upon establishing a consistent criterion for their sharpness ratings.

3. Results and discussion

3.1 Sharpness matching results

Results of the sharpness matching experiment showed that an increase in high-frequency energy must be balanced by an increase in low-frequency energy to maintain equal perceived sharpness. The open circle plotted in the upper left corner of the graph shown in Fig. 2 (left) provides the extreme example for this conclusion. When the high-frequency shelving filter had a pole frequency fixed at 8.0 kHz, the median frequency set for the low-frequency shelving filter that produced matching perceived sharpness was 3.6 kHz. This combination of values corresponds to the least flat spectral envelope shown in Fig. 2 (right), in which a low-frequency boost of around 8 dB is required to balance a high-frequency boost of around 12 dB. All ten of the spectral envelopes shown in that graph produced sharpness percepts that matched that of the spectrally flat standard stimulus. Note that the ten stimuli producing equal perceived sharpness

did not yield equal calculated values of ZS. In order to confirm this mismatch between predicted and reported sharpness, stimuli exhibiting a wider range of ZS values were presented in a sharpness rating task.

3.2 Sharpness rating results

Direct ratings of perceived sharpness were averaged across three subjects and 15 repetitions to give the 20 mean sharpness ratings plotted in Fig. 2 (left). The abscissa shows the calculated ZS values, and the sharpness ratings for each of the 20 stimuli were normalized relative to the maximum extent of the slider used in making responses. It is apparent that the 20 mean ratings of the sharpness of these stimuli are clustered around four rating magnitudes, and these four groups of stimuli are distinguished by the multiplication factors applied to the pole frequencies of the shelving filter producing low frequency boost. Though the high pole frequency increased from 8.0 to 13.9 kHz within each of the four groups of stimuli, the perceived sharpness of stimuli within each group was kept relatively constant by concurrently lowering the low shelving filter pole frequencies. So while ZS values increased with the increasing high frequency boost, perceived sharpness did not increase accordingly.

Of course, this rating experiment was designed to confirm this mismatch between predicted and reported sharpness, and the stimuli were specifically selected to reveal this failure of calculated values of ZS in predicting perceived sharpness of broadband noise. Nonetheless, it is of real value to determine how this conventional sharpness prediction equation might be modified to provide a better fit to the obtained data. It was hypothesized that a single additional term might be added to the prediction equation that would be sensitive to the stimulus parameter that keeps the predicted sharpness value constant as the first moment, or spectral centroid, of the stimulus increases. Since in the current experiment the manipulated stimulus parameter was the spread of the spectral energy, or spectral variance, the most natural choice is to add the second moment of the specific loudness distribution to the first moment. But in order to determine whether the second moment about the mean was truly the best additional term to include, the third and fourth standardized moments, skewness and kurtosis, were also considered, along with the product of two terms to capture the predictive potential of including in the equation the interaction between ZS and spectral variance.

3.3 An improved sharpness predictor

In an attempt to quantitatively model the results found here, five potential predictors of reported sharpness were submitted to stepwise multiple regression analysis, using the obtained 900 ratings of sharpness for 20 stimuli as the dependent variable. The stepwise analysis began with a model equation using only ZS as the sole predictor, with $R^2=0.767$. The next term added to the equation was the product of ZS and spectral variance, the inclusion of which caused a significant increase in R^2. The stimulus spectral variance was calculated as

$$s^2 = \frac{\sum_{z=0}^{24\ \text{bark}} (N' - \mu_{N'})^2 \cdot z}{\sum_{z=0}^{24\ \text{bark}} N'}. \tag{2}$$

The integral in the denominator represents the total loudness of the sound, and "Σ" was used rather than "∫" to indicate the finite sum of specific loudness N'.

The stepwise regression allowed for the exclusion of the remaining three predictors, variance, skewness, and kurtosis, with partial correlation coefficients of 0.185, 0.058, and 0.059, respectively. The partial correlation between reported sharpness and ZS, when controlling for the interaction term, was 0.658, whereas the partial correlation between reported sharpness and the interaction term, when controlling for ZS, was −0.418. Both of these are significant at $p < 0.001$.

The resulting improvement in the prediction of obtained sharpness ratings is apparent in the right panel of Fig. 2. The abscissa shows predicted sharpness values using the developed prediction equation. The ordinate plots for each of the 20 experimental stimuli the same mean

sharpness ratings that were plotted in the figure's left-panel graph. The goodness of the fit in these two cases was assessed using the coefficient of determination R^2 that expresses the proportion of the variance for which the prediction equation can account. Although there is a relatively large amount of error variance observed in the sharpness ratings, each of the two predictors gives a reasonably good fit: $R^2=0.767$ for ZS and $R^2=0.807$ for the two-term sharpness predictor. However, an examination of the residuals for the ZS regression reveals considerably more systematic structure that in the case of the two-term predictor. This is a natural consequence of the "tightening up" of the predictions around the mean sharpness ratings for each of the four groups of stimuli, the grand means for which were 0.14, 0.33, 0.63, and 0.83 (correlated at $r=0.98$ with the low-pole-frequency multiplication factors of 0.44, 0.67, 1.00, and 1.50).

No attempt was made here to transform the mean sharpness ratings from the observed domain of "normalized slider position" into the acum units predicted by the ZS equation. Suffice it to say that the grand mean of 900 ratings, which was 0.48, should correspond to the mean ZS of the 20 experimental stimuli, which was roughly 2.1 acum.

4. Conclusion

Two experiments were executed to observe variation in the perceived sharpness of synthesized noise stimuli having broadband spectra that were shaped by a combination of two shelving filters, one providing a low-frequency boost and the other providing a high-frequency boost. Simultaneously boosting high and low frequencies of these broadband noise stimuli for a sharpness matching task, a set of five stimuli were found that were matched in perceived sharpness, but were observed to vary relatively widely in measured ZS. In order to determine how good the ZS prediction might be for a wider range of spectral variation, sharpness ratings were made for 20 selected stimuli varying in predicted ZS from 1.6 to 2.6. By itself, ZS was shown to predict substantial variation in perceived sharpness of the 20 broadband noise stimuli. However, an examination of the residuals for the regression of reported sharpness on ZS revealed considerable systematic structure. As only a weighted spectral centroid, ZS was insensitive to manipulations in stimulus spectral variance that influenced perceived sharpness. An improved sharpness prediction equation was found via stepwise multiple regression analysis, which added a second term to the equation, a term that captured the interaction between the weighted first moment and the second moment of the stimulus specific loudness distribution. This improved sharpness prediction equation represents an advance that may begin to bridge the gap between sharpness prediction for synthetic narrow-band noise stimuli and sound stimuli with more arbitrary spectra.

References and links

[1]Eberhard Zwicker and Hugo Fastl, (**1999**), *Psychoacoustics*, Springer Series in Information Sciences, Vol. **22**, 2nd ed. (Springer-Verlag, Berlin).
[2]William L. Martens and Atsushi Marui, (**2003**), "Psychophysical calibration of sharpness for multiparameter distortion effects processing," in *Proceedings of 114th Convention*, Amsterdam, The Netherlands March 2003 (Audio Engineering Society, 2003), Preprint 5739.
[3]G. von Bismarck, (1974), "Timbre of steady sounds: A factorial investigation of its verbal attributes," Acustica **30**, 146–159.
[4]G. von Bismarck, (**1974**), "Sharpness as an attribute of the timbre of steady sounds," Acustica **30**, 159–172.
[5]Jeremy Marozeau, Alain de Cheveigné, Stephen McAdams, and Suzanne Winsberg, (**2003**), "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**(5), 2946–2957.
[6]Stephen McAdams, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete, and Jochen Krimphoff, (**1995**), "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**(3), 177–192.
[7]Malcolm Slaney, (**1994**), "Auditory toolbox: A matlab toolbox for auditory modeling work," Technical Report 45, Apple Computer (Apple Corporate Library, Cupertino, CA. New York).
[8]William Morris Hartmann, (**1997**), *Signals, Sound, and Sensation* (AIP Press, New York).
[9]William L. Martens and Atsushi Marui, (**2004**), "Predicting perceived sharpness," http://www.music.mcgill.ca/~wlm/research/sharpness/ Website created June 9, 2004.

Directionality of sperm whale sonar clicks and its relation to piston radiation theory

Kristian Beedholm
Institute of Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
beedholm@mail.dk

Bertel Møhl
Institute of Biology, University of Århus, Building 131, 8000 Århus C, Denmark
bertel.moehl@biology.au.dk

Abstract: This paper investigates the applicability to sperm whales of the theory of sound radiating from a piston. The theory is applied to a physical model and to a series of sperm whale clicks. Results show that wave forms of off-axis signals can be reproduced by convolving an on-axis signal with the spatial impulse response of a piston. The angle of a recorded click can be estimated as the angle producing the spatial impulse response that gives the best match with the observation when convolved with the on-axis wave form. It is concluded that piston theory applies to sperm whale sonar click emission.

PACS numbers: 43.80.Ka, 43.64.Tk [CFM]
Date Received: August 31, 2005 **Date Accepted:** November 28, 2005

1. Introduction

Odontocete whales use directional clicks for echolocation.[1] The clicks are generated in the nasal complex below the blowhole and transmitted via the melon to the water.[2] In dolphins where this has been investigated, it has been shown that the directional pattern of the beam can be modeled to a first approximation if one assumes the sound generator to be a flat circular piston.[3] The question is whether this model can be applied to the sperm whale that has a radically hypertrophied sound-generating nasal complex.[2]

In recordings of free-ranging sperm whales, there is no simple way to determine the relative orientation of the animal with respect to the hydrophone. The recording aspect to a piston transducer not only influences the recorded peak amplitude of the transmitted signal relative to the on-axis signal, but it also affects the wave form in a predictable way.[4] If the piston model is in fact applicable to sperm whales, it should be possible to determine the angle to the acoustic axis of the animal by comparing the observed wave forms in a scan of clicks with theoretical ones obtained from two parameters, namely the diameter of the piston and the wave form of the on-axis signal. The observed amplitudes can then be plotted against the estimated angles and compared with the predictions from the piston model.

To address the applicability of this method, we evaluated the response of an experimental physical piston, namely an electrostatic loudspeaker,[5] and compared the results to a sequence of sperm whale clicks (Fig. 1). The speaker is known to behave roughly like a piston, but has some features in common with the sperm whales, such as the lack of a baffle and presumably also some effects of the edge being less active than the center. We therefore assume that if the proposed method performed satisfactorily with the speaker, it would be likely to work with the sperm whales too.

2. Materials and methods

The theory describing the directionality of a piston transducer is well established. In textbooks, one finds that the radiation pattern of a circular baffled piston transmitting a given sound can be

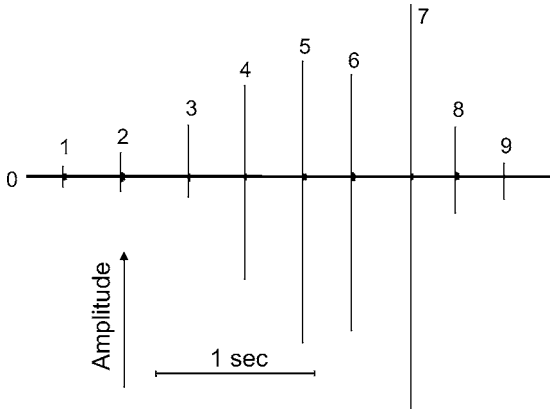


Fig. 1. Example of a scan. Sperm whale clicks from a single animal (48 kHz sampling rate) recorded at Andenes, Norway (see Ref. 4).

computed by modifying the complex spectrum of the sound by multiplication with a so-called Jinc function.[7] This spectral approach is cumbersome to derive and it is demanding to visualize the physical reasoning behind it. In contrast, it is quite straightforward to grasp the angle-dependent wave form of a piston transmitting an impulse in the time domain. In agreement with Harris[8] (citing Rutgers),[9] we shall denote this the "spatial impulse response." The appearance of any transmitted sound at a given observation angle is found by convolving the angle-dependent spatial impulse response of the piston with the sound that was observed on axis.

The two approaches, the multiplication of the complex spectrum with a Jinc function and the convolution in the time domain with the corresponding spatial impulse response, give exactly the same results. Here, we use the time domain representation, which makes it simpler to comprehend the geometry of the process.

In the direction perpendicular to the surface of a circular piston, the spatial impulse response is a theoretical Dirac function, since (per definition) it does not alter the transmitted sound through convolution. The impulse response obtained at any other angle takes the shape of a half-ellipse when graphed. To recognize this, one must accept that the pulse projected from different parts of the piston surface arrives at the receiver with different delays: The first part to arrive was projected by the closest edge, the last part of the received signal by the edge farthest away. The delay between these two time instances is (see Fig. 2)

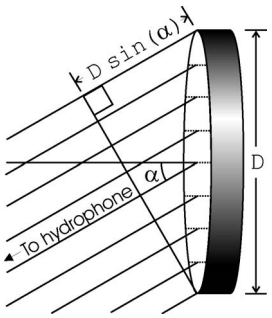


Fig. 2. Geometry of a piston transducer radiating sound as seen at a long distance relative to the diameter of the surface, illustrating the reasoning behind the spatial impulse response of a circular piston. All points on the surface radiate as a point source. If the radiated sound is an infinitely short, positive-going pulse, then the duration of the received sound becomes $\sin(\alpha)D/c$. During this time, the amplitude of the recorded sound is given by the length of the cross section (dotted lines on the surface of the disk) that projects the pulse at any one moment, leading to the sound resembling an ellipse when graphed.

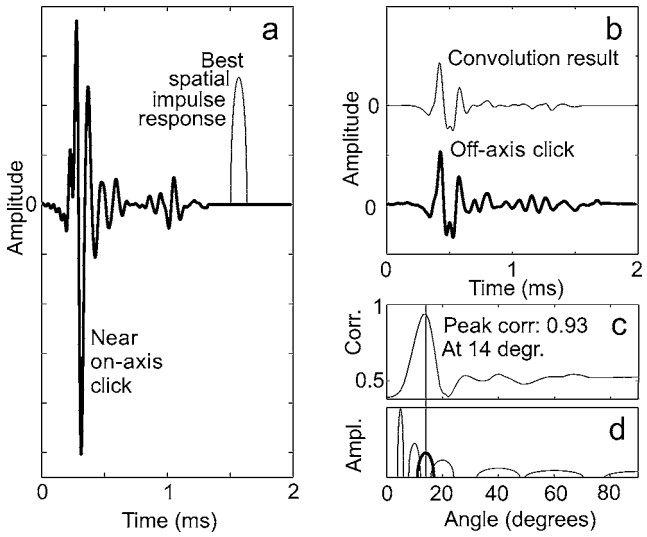


Fig. 3. Convolution and matching method for a sperm whale click. The two clicks in this example are from a scan other than the one shown in Fig. 1. (a) *Thick line* is a click judged to be recorded on axis. *Thin line* is the spatial impulse response of a piston at 14° (assuming $\varnothing$=0.8 m). (b) *Thin line* is the convolution of the two wave forms in (a). *Thick line* is the observed off-axis click to be matched. Vertical scale is the same as in (a). (c) Correlation values between the observed click in (b) and spatial impulse responses of a piston at different angles. (d) Spatial impulse response of a piston ($\varnothing$=0.8 m) in salt water at various angles. Time axis for these spatial impulse responses is the same as in (a) and (b). The spatial impulse response shown with *thick line* is the same as the thin line in (a).

$$T = \sin(\alpha)\frac{D}{c}. \tag{1}$$

During this time interval, the instantaneous cross section of the piston to transmit the sound is described as the edge-to-edge distances in a circle, measured perpendicularly to, e.g., its cosine axis. This is then the sine of the angle corresponding to the cosine value. After normalization, the resulting expression equals:

$$p(\alpha,t) = \frac{4}{\pi T}\sin\left(\cos^{-1}\left(\frac{2t}{T}\right)\right), \quad |t| < T/2, \quad 0 \text{ elsewhere}. \tag{2}$$

The factor $4/(\pi T)$ ensures that the integral is unity. Examples of such functions for different angles are given in Fig. 3(d). The expression (2) belongs to the class of functions that converge on a theoretical Dirac function as the duration goes to zero.

For both the loudspeaker and the sperm whale, $p(\alpha,t)$ was calculated for a series of closely spaced angles to make up an array of vectors representing different receiving angles. The on-axis wave form was convolved with each of these spatial impulse responses. Then, to determine the recording angle, α, of each of the (assumed) off-axis sounds, the best match was picked from the array of on-axis sounds convolved with the spatial impulse responses, $p(\alpha,t)$.

The physical model experiment was made in air with an electrostatic loudspeaker ($\varnothing$=6 cm) projecting clicks with a peak frequency of 40 kHz. The speaker was mounted on a machinist's dividing head with an angular Vernier scale. The clicks were recorded with a 0.25 in. B&K microphone (Model 4135) in a series of angles relative to the normal to the membrane. The distance[10] to the microphone was 1 m. At each angle, 256 signals were averaged after sampling with 8-bit resolution at 5 MHz.

The sperm whale click series in Fig. 1 was selected for analysis. It was recorded (16 -bit resolution, 48 kHz sampling rate)[6] at Bleijk Djup, Andenes, Norway, July 21, 2000. The whale was at a range of 1.1 km from the hydrophone and a depth of 0.64 km. The hydrophone

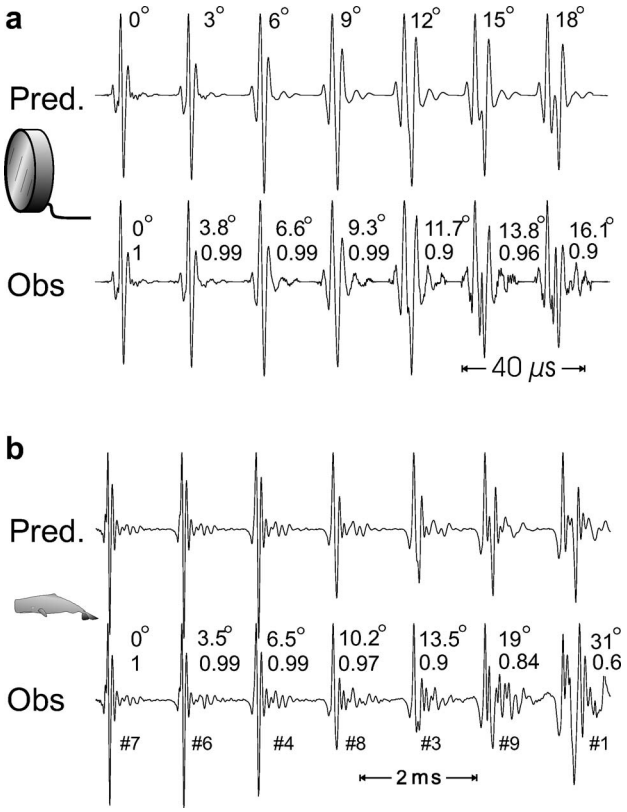


Fig. 4. Predicted and observed wave forms for an electrostatic speaker and a sperm whale. (a) Normalized click wave forms predicted and recorded from a electrostatic speaker (∅=6 cm) transmitting at the angles given above the upper traces in air. The predictions were made by convolving the wave form recorded at 0° with the relevant piston spatial impulse response. The correlation values are shown above the "observed" traces, below the angle estimated by means of the proposed method of searching for the best match between a piston transmitting the on-axis signal at different angles and the observed signal. (b) Comparison between sperm whale clicks (selected from Fig. 1) and the best matching results of submitting Click No. 7 in Fig. 1 to the piston model at different angles (with an assumed diameter of 0.8 m—see Ref. 10). The best matching angle is shown above the correlation values. Numbers below the "observed" traces indicate the pulse number in Fig. 1.

was at a depth of 30 m. From the interpulse interval of 7.4 ms, it was found[11] that the length of the animal was 16 m and, by inference, a male. Click No. 7 in Fig. 1 was considered to be on axis or nearly so.[6] It is, however, important to recognize that the degree to which a click is on axis cannot be quantified. It is further assumed that all the other clicks in the series were identical in wave form and would have had the same amplitude, had they also been recorded on axis. Since the clicks are not similar, it is conjectured that the whale is scanning with a narrow beam of sound across the general area where the hydrophone is deployed so that all clicks other than No. 7 are recorded at various angles off axis. For the theoretical predictions, the radiating aperture was tentatively set to have a cross section[12] of 0.8 m. With these assumptions, we searched for the highest correlation value between the actual observed off-axis clicks and the spatial impulse response of a piston transmitting the on-axis click [Fig. 4(b)] at 0°–90° angles (see Fig. 3) as described above. A radiation diagram was constructed for the sperm whale by plotting the peak amplitudes against the angle estimate.

3. Results and discussions

The results of the model experiment are shown in Fig. 4(a). Here, the estimated angles and the theoretically calculated wave forms are shown in the upper trace, and the known angles and the

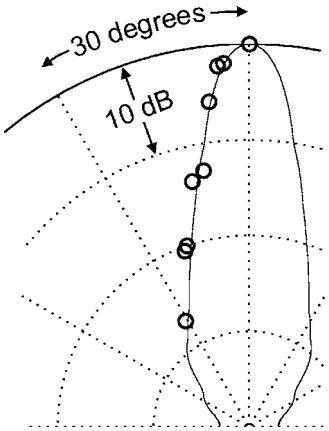


Fig. 5. Predicted and observed radiation patterns. *Thick line*: Radiation diagram of Click No. 7 in Fig. 1 (template) in a piston model with ∅=0.8 m. *Circles*: Amplitudes of the clicks in Fig. 1 plotted against the angle at which they match best with Click No. 7 as the template that was submitted to the piston model.

actually observed wave forms are shown below. The correlation between predicted and observed wave forms is generally high, highest for the smallest angles. The accuracy of the angle estimates was within 2°. This confirms that for a system known to conform to a piston model, it is possible to predict the wave form as seen off axis, using the on-axis wave form and the method proposed above. This result serves as an indicator of the way we can expect the sperm whale clicks to behave if indeed the piston model also applies to the animals. There seems to be a trend to the slight errors in the estimation of the angle, which may reflect the fact that there is no baffle and a slightly raised edge on the loudspeaker.

Figure 4(b) shows the observed and predicted wave forms for the sperm whale echolocation clicks of Fig. 1. The correlation values between them are above 0.9 out to 13.5: It therefore does appear that the observed sperm whale click wave forms are predictable as a convolution product between the on-axis click and a piston acoustic profile.

The estimated angles are on the low side in the model experiment. It is likely that this underestimation of the actual angle of recording in the case of the speaker at angles of 15° and 18° was caused by either imperfections in the behavior of the membrane or some edge effect. For the same reasons, we might suspect that the angles calculated for the sperm whales using this method will generally tend to be underestimated above, say, 12°.

Considering these caveats, the radiation diagram in Fig. 5, made by plotting the observed amplitudes of the clicks against the angle estimated with the method described in Fig. 3, does suggest that the piston model is quite applicable to these sperm whale sonar clicks, at least at these modest angles. The function in Fig. 5 (solid line) is a radiation diagram constructed from one on-axis click (Fig. 1, Click No. 7). The points are plotted at the angle of best fit for the individual click in the scan against its amplitude as obtained from Fig. 1. The degree to which the amplitude values lie on the predicted radiation diagram is independent of the computation of angle: If the sound, assumed to be recorded on axis, was in fact not so, the emerging pattern should not fit the model, and more importantly, if the model is not applicable, the beam pattern computed in this way will not comply with the prediction. The diameter of the piston does not influence how well the theoretical curve aligns with the individual amplitude points; changing the diameter changes the angles of both by an equal amount.

One interesting alternative approach to the angle estimation procedure used here is to let the loudspeaker project a scaled sperm whale click and look for the angle to the speaker where the wave form best matches the off-axis sperm whale click, in essence skipping the theoretical step. For this approach to work, however, one must deconvolve the suitably time-

compressed sperm whale click by the impulse response of the loudspeaker. In our setup, this was simply too noisy to be feasible. Several possibilities for future applications arise from the observation that the piston model seems to explain the field data.

Source levels of odontocete clicks are a much wanted datum for biosonar analysis, yet difficult to obtain in the field due to directionality and unknown orientation of the animals. While in theory it should be possible to reverse the processing outlined in Fig. 3 to obtain the on-axis signal from an off-axis representation, such an approach is confounded by zeroes in the Fourier transform of the convolution functions. However, an estimate of the on-axis amplitude may be obtained from an off-axis click by plotting the relevant piston radiation function and read out the difference in level between the estimated angle and the on-axis direction, and add this to the received level of the off-axis click.

Another promising prospect lies in collecting a larger number of sounds fulfilling the on-axis criteria.[6] It might then be possible to derive set of common characteristics (the generic sperm whale click) that allows for the determination of the off-axis angle at which any sperm whale click encountered is most likely to have been observed. With actual measured values of the angle to a sperm whale, it is possible to calibrate the success of such a general template in producing the correct angles.

Such future applications require that the diameter of the equivalent piston be determined, e.g., by recording the same click at several angles, including 0°, simultaneously. The size of the animal can be determined by applying the Gordon equation[11] to the interpulse interval of the clicks in the scan as it was done above. The piston diameter for a sperm whale of a given size can then be obtained by rearranging Eqs. (1) and (2)

Other odontocetes may be modeled in a similar way making it possible to estimate the recording angle from the characteristics of recorded clicks.

Acknowledgments

The authors thank Lee A. Miller for providing the electrostatic loudspeaker, as well as for help with the manuscript. They also thank Walther Zimmer and two anonymous reviewers for valuable comments on earlier drafts of the manuscript.

References and links

[1]W. W. L. Au, P. W. B. Moore, and D. Pawloski, "Echolocation transmitting beam of the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **80**, 688–691 (1986).
[2]T. W. Cranford, M. Amundin, and K. S. Norris, "Functional morphology and homology in the odontocete nasal complex: Implications for sound generation," J. Morphol. **228**, 223–285 (1996).
[3]P. M. Morse, *Vibration and Sound* (McGraw–Hill, New York, 1948), Chap. 7.
[4]W. A. Au, *The Sonar of Dolphins* (Springer, New York, 1993), Chap. 6.5.
[5]W. Kuhl, G. R. Schodder, and F. K. Schröder, "Condenser transmitters and microphones with solid dielectric for airborne ultrasonics," Acustica **4**, 519 (1954).
[6]B. Møhl, M. Wahlberg, P. T. Madsen, A. Heerfordt, and A. Lund, "The monopulsed nature of sperm whale clicks," J. Acoust. Soc. Am. **114**, 1143–1154 (2003).
[7]E. W. Weisstein, *Jinc Function* ⟨http://mathworld.wolfram.com/JincFunction.htm⟩ (1999).
[8]G. R. Harris, "Review of transient field theory for baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[9]A. J. Rutgers, "Spatial impulse response of an acoustic line radiator—A study of boundary-diffraction-wave phenomena and their experimental detection," Ph.D. dissertation, Catholic University of America, Washington, D.C., 1977.
[10]With a piston diameter of 6 cm, the distance of 1 m is not sufficiently far away to justify the use of the simplified Eq. (1). Instead, an expression that does take the distances into account was used to calculate the expected wave forms in Fig. 3(e).
[11]J. C. D. Gordon, "Evaluation of a method for determining the length of sperm whales (*Physeter catodon*) from their vocalizations," J. Zool. **224**, 301–314 (1991).
[12]B. Møhl, "Sound transmission in the nose of the sperm whale *Physeter catodon*: A postmortem study," J. Comp. Physiol., A **187**, 335–340 (2001).

ACOUSTICAL NEWS—USA

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

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

Preliminary Notice: 151st Meeting of the Society of America

The 151st Meeting of the Acoustical Society of America (ASA) will be held Monday through Friday, 5–9 June 2006 at the Rhode Island Convention Center, Providence, Rhode Island. The Westin Providence, which is connected to the convention center, is the headquarters hotel. Room blocks have been reserved at The Westin Providence and an overflow hotel.

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

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 5–9 June. The special sessions described below will be organized by the ASA Technical Committees.

Special Sessions

Acoustical Oceanography (AO)

Exploitation of sound by marine mammals
(Joint with Animal Bioacoustics)
Marine mammals are sophisticated users of sound, both self-generated and the surrounding ambient. This session features research that shows examples of this behavior.
Ultralow and low-frequency marine seismo-acoustics
(Joint with Engineering Acoustics)
Focus on acoustic sensing of natural seismic phenomena such as tsunamis, earthquakes, and ocean storm events

Animal Bioacoustics (AB)

Acoustic interactions in animal groups
Behaviors and physiological mechanisms associated with the challenges and advantages of acoustic signal transmission and reception in animal groups
Effects of anthropogenic sounds on fishes
(Joint with Acoustical Oceanography and Underwater Acoustics)
Overview of issues and findings with regard to the effects of anthropogenic sound on fish and fisheries
Is hearing all cochlear?
Examine evidence for a conserved near-field otolithic acoustic sense in amniotes including humans
Mechanisms of biosonar
(Joint with Signal Processing in Acoustics and Engineering Acoustics)
How biosonar works from the perspective of field observations, laboratory experiments, and signal processing

Architectural Acoustics (AA)

Acoustics of libraries
Theory and practice of acoustical design in libraries and media centers, including study areas as well as more specialized spaces (e.g., music and film collections, multimedia rooms, etc.)
Acoustics of multifamily housing
(Joint with Noise)
Room acoustics, noise and litigation issues dealing with multifamily housing
Composed spaces
(Joint with Musical Acoustics and Engineering Acoustics)
Surround sound compositions in which space (real or created) is part of the art
Composed spaces loudspeaker concert
(Joint with Musical Acoustics and Engineering Acoustics)
Experience works by invited composers in a continuous concert in a dedicated room open to all ASA attendees
Microperforated acoustical absorbing materials
(Joint with Noise, Physical Acoustics, and Engineering Acoustics)
Theory, manufacture, and applications of microperforated acoustical absorbing materials
Spaces of worship—Another quarter century of experience (1984–present)
Poster session for spaces of worship
Special session in honor of John Kopec
Session in memory of John Kopec
Surround sound essentials
A unique opportunity to take advantage of a high-quality loudspeaker installation to investigate (and possibly give direction to) various parameters that are important in the design of "surround sound" studios and listening rooms

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Brown tadpoles and red herrings: Boiling inertial cavitation and nonlinearity in high-intensity focused ultrasound lesion formation
A forum to focus on the debate on the relative importance of nonlinearity in high-intensity focused ultrasound lesion formation
Celebration session for Edwin Carstensen
Will relate to and honor the work of Dr. Edwin L. Carstensen
Memorial session for Frederic Lizzi
Honoring the career and life of Frederic L. Lizzi
Sensing and imaging using light and sound
(Joint with Physical Acoustics and Signal Processing in Acoustics)
Devoted to sensing and imaging using various combinations of optical and acoustical energy
Targeted contrast agents
Will focus on the use of agents that target specific sites in the body and are activated by ultrasound to produce therapeutic effect

Education in Acoustics (ED)

Hands-on experiments for high school students
Twenty stations will be staffed by scientists and graduate students for high school students' hands-on experience
Medical education
Courses and programs for students and professionals in the health services
Undergraduate research poster session
Undergraduate students will present their research in poster format and will be available during the session to explain their work

Engineering Acoustics (EA)

Joe Blue memorial session part 1: Transduction, linear and nonlinear
(Joint with Physical Acoustics)
Will cover physical acoustics theory, transducers, materials, calibration, Underwater Sound Reference Division standards and their application

Musical Acoustics (MU)

Human-computer interfaces
(Joint with Engineering Acoustics and Signal Processing in Acoustics)
Design of input devices and control displays for electronic and computer music applications

 0001-4966/2006/119(2)/669/7/$22.50

Scaling of musical instrument families
Application of physical laws to aid creation of families of like instruments that cover different pitch ranges
Finite element and finite difference methods in musical acoustics
(Joint with Structural Acoustics and Vibration)
Analysis and simulation of musical instruments using finite element or finite difference methods

Noise (NS)

Audio-visual design in soundscapes
(Joint with Architectural Acoustics)
Audio-visual soundscape theory, measurement, and design practice, including field testing demonstrations
Essential acoustical curriculum for noise control techniques for engineers, industrial hygienists, and architects
(Joint with Education in Acoustics, Engineering Acoustics, and Architectural Acoustics)
Will examine what elements of noise control can be integrated into curriculum targeted to the engineer, industrial hygiene, and architecture students
Fifty years of speech privacy
(Joint with Architectural Acoustics and Speech Communication)
Celebrating the contributors to understanding and use of speech privacy methods
Heating, ventilating, and air-conditioning noise control
(Joint with Architectural Acoustics)
Theory, advances, and case studies
New loudness standard
(Joint with ASA Committee on Standards and Psychological and Physiological Acoustics)
Focused around a new (to be approved) ANSI loudness standard

Physical Acoustics (PA)

Acoustic microscopy at the nanoscale
Focused on acoustic methods in the broadest sense, including modeling and experiments, for characterization of materials with submicron length scales
Celebration of the work of Brown University
Era of R. B. Lindsay, R. T. Beyer, P. J. Westervelt, A. O. Williams, and others along with a new era in the Physics Department

Psychological and Physiological Acoustics (PP)

Characterizing auditory attention
Will summarize research teasing apart how the listener, the experimenter, and the stimuli guide how the auditory scene is processed
Individual susceptibility to noise-induced hearing loss
(Joint with ASA Committee on Standards and Noise)
Will cover aspects of noise-induced hearing loss, factors affecting susceptibility, and the current status of treatment of noise-induced hearing loss

Signal Processing in Acoustics (SP)

Adjoint modeling in acoustics
(Joint with Underwater Acoustics and Acoustical Oceanography)
Adjoint methods are used for data assimilation, model tuning, and sensitivity analysis in many fields. Brings together research in areas related to adjoint modeling, such as inverse scattering, shape reconstruction, optimal design, diffraction tomography, and ocean acoustic tomography
Auditory processing of sonar signals
(Joint with Underwater Acoustics and Psychological and Physiological Acoustics)
Various methods associated with human hearing can be applied to sonar signals. Papers in models of auditory perception, auditory effects of oceanic processes, and sonar signal processing are solicited
Optimum and robust signal processing in uncertain and random environments
(Joint with Underwater Acoustics)
New developments in signal processing are presented to perform various tasks in inhomogeneous and/or time-varying environments
Processing of acoustic vector sensors
(Joint with Underwater Acoustics and Engineering Acoustics)
Realized benefits of vector sensors in conjunction with pressure sensors that enable measurements of intensity include reduced aperture requirements, better spatial performance, etc. Modeling and measurements of sound intensity and signal processing that exploit properties of noise are presented
Signal processing: Diverse problems—similar solutions
Solutions developed for one physical regime are applied to another
Topics in seismic signal processing
Signal processing of various seismic signals is discussed

Speech Communication (SC)

Enhancement and multiplicity of cues in speech
Perceptual cues in speech and relationship between enhancement and multiplicity of cues

Structural Acoustics and Vibration (SA)

Ultrasonic waveguides for structural monitoring
The use of ultrasonic waveguides in structures to explore properties of structures

Underwater Acoustics (UW)

High-frequency acoustic propagation and applications
The impacts of ocean processes on high-frequency underwater acoustic propagation and their application to tomography and acoustic communications
High-frequency ambient noise
(Joint with Acoustical Oceanography)
Measurements and modeling comparisons of ambient noise with emphasis on mid-to-high frequencies and anisotropy
Joe Blue memorial session part 2: Shallow water and marine animal acoustics
(Joint with Acoustical Oceanography and Animal Bioacoustics)
Will cover propagation, marine animal acoustics, noise, and masking
Scattering of sound at the sea surface
Will focus on quantifying the interaction of sound waves with a time-varying sea surface on all scales of roughness. Contributions involving modeling and/or measurements (field or laboratory) are encouraged

Other Technical Events

Exhibit

The meeting will be highlighted by an exhibit which will feature displays with instruments, materials, and services for the acoustical and vibration community. The exhibit, which will be conveniently located near the registration area and meeting rooms, will open at the Rhode Island Convention Center with a reception on Monday evening, 5 June, and will close Wednesday, 7 June. Morning and afternoon refreshments will be available in the exhibit area.

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

Technical Tour

A tour of the acoustic test facilities at the Naval Undersea Warfare Center (NUWC) in Newport, Rhode Island will be conducted on Monday, 5 June 2006. These facilities include: acoustic tank, pressure tank, antenna test chamber, and anechoic chamber. Lunch will be served in the Officer's Club immediately following the technical tour. The bus will leave the Rhode Island Convention Center at 9:00 a.m. and return at approximately 3:00 p.m. The Technical Tour fee is $35.00. To register please e-mail asaprov06@cox.net. A NUWC visitor request form will be forwarded to you and MUST be on file prior to the tour. Deadline for registration is 1 March 2006.

Distinguished Lecture

A Distinguished Lecture will be presented by Nikolai Andreevich Dubrovsky, Director of the N. N. Andreyev Acoustics Institute, Russian Academy of Science and President of the Russian Acoustical Society.

Hot Topics Session

A "Hot Topics" session sponsored by the Tutorials Committee is scheduled covering the fields of Acoustical Oceanography, Education in Acoustics, and Underwater Acoustics.

Student Design Competition

The Technical Committee on Architectural Acoustics of the Acoustical Society of America and the National Council of Acoustical Consultants are sponsoring a Student Design Competition to be displayed and professionally judged at the Providence meeting. The 2006 competition involves the design of a City Municipal Building, including a Council Chambers and Courtroom. Individual students or teams of a maximum of three may submit entries. Graduate and undergraduate entries are welcome. Attendance at the Providence meeting is not required for entry or award in the competition. Submissions will be poster presentations that demonstrate room acoustics, noise control, and acoustic isolation techniques in building planning and room design.

An award of $1,250 US will be made to the submitter(s) of the entry judged "First Honors." Four awards of $700 US each will be made to submitters of entries judged "Commendation."

Registration deadline is 3 April 2006. Full details about registration, the competition, and the design scenario are available at ⟨www.newmanfund.org⟩ or can be requested by contacting Norm Phillip; (206) 224.3676; normp@yantis.com

Gallery of Acoustics

The Technical Committee on Signal Processing in Acoustics will sponsor the eighth Gallery of Acoustics at the Providence meeting. The objective of the Gallery is to enhance ASA meetings by providing a setting for researchers to display their work to all meeting attendees in a forum emphasizing the diversity, interdisciplinary, and artistic nature of acoustics.

The Gallery will consist of a multimedia collection of images, videos, audio clips, and narrations, of images and/or sounds generated by acoustic processes or resulting from signal and image processing of acoustic data. Images and videos can consist of actual visualizations of acoustic processes, or of aesthetically and technically interesting images resulting from various signal and image processing techniques and data visualization. Audio clips and segments should also have both aesthetic and technical appeal.

Entries must be submitted electronically, either by e-mail attachment, CD, or DVD. A panel of referees will judge entries on the basis of aesthetic/artistic appeal, ability to convey and exchange information, and originality. A cash prize of $350 will be awarded to the winning entry. The top three entries will be posted on the Gallery web site: www.sao.nrc.ca/ims/asa_sp/Gallery.html

Further details including submission formats, instructions, and deadlines can be found in the printed call for papers or online at ⟨http://asa.aip.org/providence/providence.html⟩.

Grant Writing Workshop

The Student Council is pleased to announce a workshop on grant writing for student and postdoctoral members of the ASA to be offered during the Providence meeting. The workshop will focus on the mechanics of grant writing for members of all technical committees. The specific topics that will be covered will include: white papers and letter proposals; full proposals; essential components; budget writing; common mistakes; what reviewers look for; rejected? what next? Examples of pre/post-revision funded proposals will be made available on the Student Council website. These proposals will be discussed during the workshop. The duration of the workshop is 1.5 hours. Look for specific time/date information on the website for students (http://www.acosoc.org/student/), in the student E-zine, the Student Guide, and on the Student Council bulletin board at the meeting.

Online Meeting Papers Website

The ASA has replaced its traditional at-meeting "Paper Copying Service" with an online site which can be found at ⟨http://scitation.aip.org/asameetingpapers/⟩. Authors of papers to be presented at meetings will be able to post their full papers or presentation materials for others who are interested in obtaining detailed information about meeting presentations. The online site will be open for author submissions in April. Submission procedures and password information will be mailed to authors with the acceptance notices.

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

Meeting Program

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

Tutorial Lecture

A tutorial presentation on The 2004 Sumatra earthquake and tsunami: Multidisciplinary lessons from an oceanic monster will be given by Emile A. Okal of Northwestern University on Monday, 5 June at 7:00 p.m.

Lecture notes will be available at the meeting in limited supply. Those who register by 15 May are guaranteed receipt of a set of notes.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15 for registration received by 8 May and $25 thereafter, including on-site registration at the meeting. The fee for students with current ID cards is $7 for registration received by 8 May and $12 thereafter, including on-site registration at the meeting. Use the registration form in the printed call for papers or register online at ⟨http://asa.aip.org⟩.

Short Course on Underwater Acoustic Communications

Underwater acoustic telemetry is the art of transmitting information between distant points in the ocean, rivers, or lakes. Some applications require only small transmission rates, such as the command and control systems for ocean equipment in the offshore industry and acoustic releases for oceanographic moorings.

The objective of this course is to provide a basic understanding of underwater acoustic telemetry in deep and shallow water. A review of underwater sound propagation in deep and shallow water, and modeling of underwater acoustic communication systems, will be provided. State-of-the-art developments in equipment, signal processing, and networking techniques will be presented. Examples and experimental results will be presented throughout the course.

The instructor is Pierre-Philippe Beaujean, Assistant Professor in the Department of Ocean Engineering at Florida Atlantic University. Dr. Beaujean leads the Acoustic Communications and Navigation Research Laboratory within the Center for Acoustics and Vibrations at Florida Atlantic University.

The course will be held on Sunday, 4 June 2006, 1:00 p.m. to 6:00 p.m. and Monday, 5 June 2006, 8:30 a.m. to 12:30 p.m.

The registration fee is $250 and covers attendance, instructional materials, and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. Only those who have registered by 15 May will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 8 May. Full refunds will be made for cancellations prior to 8 May. Any cancellation after 8 May will be charged a $25 processing fee. Use the registration form in the printed call for papers to register or register online at ⟨http://asa.aip.org⟩.

Special Meeting Features

Student Transportation Subsidies

A student transportation subsidies fund has been established to provide limited funds to students to partially defray transportation expenses to meetings. Students presenting papers who propose to travel in groups using economical ground transportation will be given first priority to receive subsidies, although these conditions are not mandatory. No reimbursement is intended for the cost of food or housing. The amount granted each student

depends on the number of requests received. To apply for a subsidy, submit a proposal (e-mail preferred) to be received by 8 May to: Jolene Ehl, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359, Fax: 516-576-2377, e-mail: jehl@aip.org The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto, whether you will travel alone or with other students; names of those traveling with you; and approximate cost of transportation.

Young Investigator Travel Grant

The Committee on Women in Acoustics is sponsoring a Young Investigator Travel Grant to help with travel costs associated with presenting a paper at the Providence meeting. This award is designed for young professionals who have completed the doctorate in the past five years (not currently enrolled as a student), who plan to present a paper at the Providence meeting. Each award will be of the order of $300. It is anticipated that the Committee will grant a maximum of three awards. Applicants should submit a request for support, a copy of the abstract they have submitted for the meeting, and a current resume/vita which provides information on their involvement in the field of acoustics and to the ASA to: Dr. Donna L. Neff, Boys Town National Research Hospital, 555 North 30th Street, Omaha NE 68131; Tel.: 402-452-5069; Fax: 402-452-5027; e-mail: neff@boystown.org Deadline for receipt of applications is 28 April.

Students Meet Members for Lunch

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

Plenary Session, Awards Ceremony, Fellows Luncheon, and Social Events

Buffet socials with cash bar will be held on Tuesday and Thursday evenings at the Rhode Island Convention Center.

The ASA Plenary session will be held on Wednesday afternoon, 7 June at the Rhode Island Convention Center, where Society awards will be presented and recognition of newly elected Fellows will be announced.

A Fellows' Luncheon will be held on Thursday, 8 June at 12:00 noon. This luncheon is open to all attendees and their guests (you need not be a Fellow to attend). To purchase tickets, use the registration form in the printed call for papers or the online form at ⟨http://asa.aip.org⟩.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 7 June. Those who wish to attend this luncheon must register using the form in the printed call for papers or online at ⟨http://asa.aip.org⟩. The fee is $15 (students $5) for preregistration by 8 May and $20 (students $5) thereafter, including on-site registration at the meeting.

Special Event

There will be an open exhibit of China Blue's work on Thursday, 8 June 2006 from 4:00 p.m. to 6:00 p.m. China Blue is an internationally known sound artist from New York. She is interested in how sound defines and articulates a space. Her work uses samples of acoustic events from day-to-day life as a way of examining and teaching about sonic flow and energy. Her spatialized recording techniques are based on the psychophysics underlying both auditory and visual localization. For additional information please contact Andrea Simmons, andrea_simmons@brown.edu

Transportation and Hotel Accommodations

Air Transportation

Providence is served by T. F. Green Airport (Airport Code PVD). Only ten minutes from downtown Providence, T. F. Green Airport was recently named one of the "Top Five Alternative Airports" in the country by Forbes magazine, and received a Reader's Choice recommendation from Condé Naste Traveler. Just off Exit 13 on Interstate Route 95, T. F. Green Airport is accessible to Boston, Cape Cod, and Southeastern New England, and is fast becoming a popular alternative to Boston's Logan Airport. T. F. Green Airport offers more than 160 direct flights via major carriers such as American, Continental, Delta, Northwest, Southwest, Spirit, United, US Air, and many others. For flight information visit, ⟨www.pvd-ri.com⟩; for other information of interest visit ⟨www.ProvidenceNightandDay.com⟩.

Ground Transportation

Transportation from the T. F. Green Airport to The Westin Providence and the overflow hotel:

Information. The Information Booth, located in the baggage claim area, serves many needs of the traveler to T. F. Green Airport, including Courtesy Paging, Parking, and Directions to locations in the local area. The staff can also provide general information on what is at T. F. Green, such as ATMs, telephones, etc. The information booth is staffed from 6:00 a.m. to 11:00 p.m.

Rail Service. Providence is located on Amtrak's Northeast Corridor between Washington DC/New York City and Boston. High-speed Acela Express train service transports passengers form New York City to Providence in about two and a half hours. The Massachusetts Bay Transit Authority (MBTA) runs low-cost commuter trains to Providence from Boston and other points in Massachusetts. Amtrak's Providence railway station is within walking distance of The Westin Providence, The Rhode Island Convention Center, and the overflow hotel. Amtrak's contact information: 1-800-USA-RAIL or visit ⟨www.amtrak.com⟩.

Major car rental companies. Nearly every major car rental company is represented at T. F. Green. Rental car counters are located adjacent to the baggage claim area. You may also access a particular company online using T. F. Green Airport's ground transportation link ⟨www.pvd-ri.com/ground_transport/car_rentals.htm⟩.

Parking at The Westin Providence. The hotel self-parking rate is $20 per day; valet parking rate is also $20 per day.

Parking at the overflow hotel. The hotel self-parking rate is $19 per day; no valet service. Airport Limousine Shuttle Service, shared-ride, door-to-door service. The shuttle departs T. F. Green Airport every hour on the hour. It arrives at the overflow hotel at 15 minutes past every hour and at The Westin Providence at 17 minutes past every hour. The fee for this service is $9 per person each way. Please note: After 7:00 p.m. reservations for the shuttle are required from the city only. Baggage fee: Everyone is allowed 2 pieces of baggage. Anything in excess of the 2 pieces incurs a $2 fee per bag. Airport Limousine Shuttle runs daily from 5:00 a.m. to 11:00 p.m. Phone 401-737-2868 or visit: ⟨www.airporttaxiri.com⟩.

Taxicabs and limousines. Taxis are available outside the terminal at T. F. Green Airport. Providence is approximately 10 minutes from the airport, with fares averaging $35 one way. All cab fares are metered. Please phone 401-737-2868 or Toll-Free 1-888-737-7006 for more information.

Driving information. Located at the intersection of I-95 and I-195, Providence is 50 miles from Boston (about a one-hour car ride) and 185 miles from New York City (about three hours by car).

Hotel Reservation Information

Registration, technical sessions, and socials will be held at The Rhode Island Convention Center. Administrative meetings will be held at the headquarters hotel, The Westin Providence, which is connected to the convention center. Blocks of sleeping rooms are available at The Westin Providence and an overflow hotel. Early reservations are strongly recommended. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America (ASA). The reservation cutoff date for the special discounted ASA rates is 3 May 2006; after this date, the conference rates will no longer be available.

The Westin Providence, which is the headquarters hotel, is located in the heart of downtown and is connected to The Rhode Island Convention Center and The Providence Place Mall.

The hotel features a fully equipped health club, indoor heated swimming pool, sauna, jacuzzi, as well as a full service business center. For more details visit ⟨www.westin.com/providence⟩.

Please make your reservation directly with The Westin Providence. When making your reservation, you must mention the Acoustical Society of America to obtain the special ASA meeting rates.

The Westin Providence
One West Exchange Street
Providence, RI 02903
Tel.: 401-598-8000; Toll-Free: 1-800-368-7764
Fax: 401-598-8200
Online: www.starwoodmeeting.com./StarGroupsWeb/res?id=0511083475&key=B87A0
Rates:
Single/Double: $164 plus tax
Reservation cutoff date: 3 May 2006

OVERFLOW HOTEL
THE COURTYARD PROVIDENCE DOWNTOWN
The Courtyard Providence Downtown
32 Exchange Terrace at Memorial Blvd.
Providence, RI 02903
Tel.: 401-272-1191; Toll-Free: 1-888-887-7955
Fax: 401-272-1416
Online: www.marriott.com/property/propertypage/PVDDT
king bed, code-ACSACSA
two queen beds, code-ACSACSB
Rates:
Single/Double $149 plus tax
Reservation cutoff date: 3 May 2006

Room Sharing

ASA will compile a list of those who wish to share a hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, by 8 May to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Acoustical Society of America, Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

June is one of the most beautiful months in Rhode Island, when spring is just giving way to summer. Let the enjoyable warm temperatures start your day and light summer breezes follow you into evening. Although June is typically quite comfortable, sunscreen and a hat are suggested when enjoying outdoor activities. Rhode Island's famous beaches are open for the season and located just a short distance from Providence. For those who prefer an air-conditioned environment, The Providence Place Mall is connected by walkway to The Westin Providence. Average high temperature in June is 77 degrees F, with average lows around 58 degrees F. Average precipitation is 3.38 inches. For additional information on Rhode Island weather, visit: ⟨iwin.nws.noaa.gov/iwin/ri/ri.html⟩.

Assistive Listening Devices

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

Accompanying Persons Program

Spouses and other visitors are welcome at the Providence meeting. The registration fee for accompanying persons is $50 for preregistration by 8 May and $75 thereafter, including on-site registration at the meeting. A hospitality room for accompanying persons will be open in The Westin Providence's Blackstone Room from 8:00 a.m. to 5:00 p.m. each day throughout the meeting, where information about activities in and around Providence will be provided.

Rhode Island's small size is a big advantage for you. The state is a convenient 37 miles wide by 48 miles long, which means that legendary mansions, scenic beaches, and fascinating historical sites are all with minutes of Providence. Other popular destinations like Cape Cod and Mystic Seaport are less than an hour away. We are also within easy traveling distance of several major northeast cities: Boston—50 miles, New York City—180 miles, Montreal—under 350 miles. Providence is so close to everything, you'll be able to pack a lot of fun into a short amount of time. Make sure to visit each of Rhode Island's wonderful regions:

Blackstone Valley—Explore fascinating places like Slater Mill, birthplace of the Industrial Revolution. Tour the Blackstone Valley National Heritage Corridor. Browse the Valley's dozens of Factory Outlets. Catch some great baseball action with the Pawtucket Red Sox, Boston's Triple A farm team. Or, get the true small-town New England experience at the Brown and Hopkins General Store. It's the most enjoyable history lesson you'll ever have. Visit ⟨www.tourblackstone.com⟩.

Block Island—Block Island, an 11-square-mile seaside resort located 12 miles off the Rhode Island coast, has been heralded as "One of the Last Twelve Great Places in the Western Hemisphere." Its rolling green hills and dramatic bluffs are reminiscent of Ireland, while its beautifully restored Victorian hotels and inns preserve the elegance of a bygone era. Accessible by ferry and by air, Block Island should be part of your Rhode Island experience. Visit ⟨www.blockislandinfo.com⟩.

East Bay—The region, which is comprised of the towns of Barrington, Bristol, East Providence, and Warren, is steeped in British and Colonial heritage. Located in beautiful Bristol is Blithewold Mansion and Gardens, a waterfront estate which is home to thousands of flowers. Enjoy a bay cruise or some antiquing on Water Street in Warren. Museums like Coggeshall Farm and the Herreshoff Marine Museum provide a look at Rhode Island's agricultural and nautical past. Visit ⟨www.eastbayritourism.com⟩.

Newport—You'll find the summer "cottages" of some of the world's wealthiest people here in the City-by-the-Sea. These world-renowned Gilded Age mansions on famous Bellevue Avenue line Newport's rugged coastline. Home of America's Cup yacht racing and the Newport Jazz Festival, this remarkable city has gained international fame amongst pleasure-seeking travelers. Visit ⟨www.gonewport.com⟩.

Providence—The city is acclaimed for its preservation of the past and its advancements toward the future. A true Renaissance City. Places like Benefit Street's "Mile of History" on the East Side to festive Federal Hill, Rhode Island's own "Little Italy," feature historical and cultural attractions. The Providence Place Mall, one of the largest shopping destinations in the country, is a Downcity favorite. Providence is home to Barnaby Evans' award-winning river sculpture, WaterFire. Rhode Island's capital is also endowed with exceptional arts and entertainment, famed museums, world-class restaurants, and a nightlife that continues to grow.
Visit ⟨www.goprovidence.com⟩.

Warwick—As Rhode Island's retail shopping mecca, the City of Warwick attracts hundreds of thousands of eager spenders every year. Hundreds of specialty shops and discount stores, plus two malls housing major department stores, famous retail shops, and one-of-a-kind boutiques line the city's stretch of Route 2, a four-lane thoroughfare which has become known as Rhode Island's "Miracle Mile of Shopping."
Visit ⟨www.visitwarwickri.com⟩.

South County—Come relax on miles of unspoiled beaches. Sunning, swimming, golfing, fishing, and whale watching are just a few of the many seaside activities to be enjoyed. Charming New England towns and historic sites like the South County Museum, Gilbert Stuart's Birthplace, and the Indian Cultural Center provide stimulating glimpses of the past. Watch the trawlers come and go from the breakwater in the working fishing village of Point Judith, town of Narragansett. Visit ⟨www.southcountyri.com⟩.

Registration Information

The registration desk at the meeting will open on Monday, 5 June, at the Rhode Island Convention Center. To register use the form on page 17 or register online at ⟨http://asa.aip.org⟩. If your registration is not received at the ASA headquarters by 15 May you must register on-site.

Registration fees are as follows:

Category	Preregistration by 8 May	Registration after 8 May
Acoustical Society Members	$350	$425
Acoustical Society Members One-Day	$175	$215
Nonmembers	$400	$475
Nonmembers One-Day	$200	$240
Nonmember Invited Speakers One-Day Only	Fee waived	Fee waived
Nonmember Invited Speakers (Includes one-year ASA membership upon completion of an ASA application)	$110	$110
ASA Student Members (with current ID cards)	Fee waived	Fee waived
Nonmember Students (with current ID cards)	$40	$50
Emeritus members of ASA	$50	$75
(Emeritus status preapproved by ASA)		
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$50	$75

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2006) of membership. Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** may register for one day only without charge. A nonmember invited speaker who pays the full-week registration fee will be given one free year of membership upon completion of an ASA application form.

Note: A $25 fee will be charged to those who wish to cancel their registration after 8 May.

Online Registration

Online registration is now available at ⟨http://asa.aip.org⟩.

Members of the Local Committee for the Meeting

General Chair: James H. Miller; Technical Program Chair: James F. Lynch; Food Service/Social Events/Meeting Administrator: Gail Paolino; Audio-Visual: Peter M. Scheifele; Accompanying Persons Program/Cultural Attaché: James A. Simmons and Andrea M. Simmons; Signs/Publicity: John R. Buck; Technical Tour: David Moretti; Meeting Room Coordinator: Gopu R. Potty

Errata

Volume **118**(4), 2092 (2005): The article titled "Minority Fellowship Awarded to Bobby E. Gibbs" incorrectly states that Mr. Gibbs has been awarded an M.M. degree from Bowling Green State University.

USA Meetings Calendar

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

2006

16–18 Feb.	31st Annual Conference, National Hearing Conservation Association, Tampa, FL [NHCA, 7995 E. Prentice Ave., Suite 100 East, Greenwood Village, CO 80111-2710; Tel: 303-224-9022; Fax: 303-770-1614; E-mail: nhca@gwami.com; WWW: www.hearingconservation-.org].
6–9 June	151st Meeting of the Acoustical Society of America, Providence, RI [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].
17–21 Sept.	INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA [www.interspeech2006.org ⟨http://www.interspeech2006.org/⟩]
28 Nov.–2 Dec.	152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, HI [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]. Deadline for receipt of abstracts: 30 June 2006.

2007

4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, UT [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, LA (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

2008

28 July–1 Aug.	9th International Congress on Noise as a Public Health Problem (Quintennial meeting of ICBEN, the International Commission on Biological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609, Tel. 860-572-0680; Web: www.icben.org. E-mail icben2008@att.net].

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and

indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).

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

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

ACOUSTICAL NEWS—INTERNATIONAL

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

International Meetings Calendar

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

March 2006

14–16 **Spring Meeting of the Acoustical Society of Japan**, Tokyo, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

20–23 **Meeting of the German Acoustical Society (DAGA 2006)**, Braunschweig, Germany (Web: www.daga2006.de).

April 2006

3–4 *__Futures in Acoustics__, Southampton, UK (Web: www.ioa.org.uk).

24–27 **French Congress on Acoustics**, Tours, France (Web: cfa06.med.univ-tours.fr).

May 2006

2–5 **International Conference on Speech Prosody 2006**, Dresden, Germany (Web: www.ias.et.tu-dresden.de/sp2006).

5–7 **Sixth International Conference on Auditorium Acoustics**, Copenhagen, Denmark (e-mail: t.j.cox@salford.ac.uk; Web: www.ioa.org.uk/viewupcoming.asp).

10–12 *__Meeting of the Italian Acoustical Association__, Ischia (NA), Italy (Web: www.associazioneitalianadiacustica.it).

15–19 **IEEE International Conference on Acoustics, Speech, and Signal Processing**, Toulouse, France (Web: icassp2006.org).

16–19 *__IEEE Conference Oceans'06 Asia Pacific__, Singapore (Web: www.oceans06asiapacific.sg).

23–26 **17th Session of the Russian Acoustical Society**, Moscow, Russia (Web: www.akin.ru).

30–1 **6th European Conference on Noise Control (EURONOISE2006)**, Tampere, Finland (Fax: +358 9 7206 4711; Web: www.euronoise2006.org).

June 2006

12–15 **8th European Conference on Underwater Acoustics**, Carvoeiro, Portugal (Web: www.ecua2006.org).

17–19 **9th International Conference on Recent Advances in Structural Dynamics**, Southampton, UK (Web: www.isvr.soton.ac.uk/sd2006/index.htm).

26–28 **9th Western Pacific Acoustics Conference (WESPAC 9)**, Seoul, Korea (Web: wespac9.org).

26–29 **11th International Conference on Speech and Computer**, St. Petersburg, Russia (Web: www.specom.nw.ru).

July 2006

3–7 **13th International Congress on Sound and Vibration (ICSV13)**, Vienna, Austria (Web: info.tuwienac.at/icsv13).

17–20 **International Symposium for the Advancement of Boundary Layer Remote Sensing (ISARS13)**, Garmisch-Partenkirchen, Germany (Fax: +49 8821 73 573; Web: imk-ifu.fzk.de/isars).

17–19 **9th International Conference on Recent Advances in Structural Dynamics**, Southampton, UK (Web: www.isvr.soton.ac.uk/sd2006/index.htm).

September 2006

13–15 **Autumn Meeting of the Acoustical Society of Japan**, Kanazawa, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

18–20 **Sixth International Symposium on Active Noise and Vibration Control (ACTIVE2006)**, Adelaide, Australia (Web: www.active2006.com).

18–20 **International Conference on Noise and Vibration Engineering (ISMA2006)**, Leuven, Belgium (Fax: 32 16 32 29 87; Web: www.isma-isaac.be).

18–20 *__12th International Conference on Low Frequency Noise and Vibration and its Control__, Bristol, UK (Web: www.lowfrequency2006.org).

October 2006

25–28 **Fifth Iberoamerican Congress on Acoustics**, Santiago, Chile (Web: www.fia2006.cl).

November 2006

2–3 *__Swiss Acoustical Society Fall Meeting__, Luzern, Switzerland (Web: www.sga-ssa.ch).

20–22 **1st Joint Australian and New Zealand Acoustical Societies Conference**, Christchurch, New Zealand (Web: www.acoustics.org.nz).

April 2007

10–12 *__4th International Conference on Bio-Acoustics__, Loughboro, UK (Web: www.ioa.org.uk).

July 2007

9–12 **14th International Congress on Sound and Vibration (ICSV14)**, Cairns, Australia (e-mail: n.kessissoglou@unsw.edu.au).

August 2007

26–29 **inter-noise 2007**, Istanbul, Turkey (Web: www.internoise2007.org.tr).

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

September 2007

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

9–12 **ICA Satellite Symposium on Musical Acoustics (ISMA2007)**, Barcelona, Spain (SEA, Serano 144, 28006 Madrid, Spain; Web: www.isma2007.org).

9–12 *__ICA Satellite Symposium on Room Acoustics (ISRA2007)__, Sevilla, Spain (Web: www.isra2007.org).

June 2008

29–4 **Forum Acusticum: Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France (Web: www.sfa.asso.fr; e-mail: phillipe.blanc-benon@ec-lyon.fr).

 0001-4966/2006/119(2)/677/2/$22.50

July 2008

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

August 2010

23–27 **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (Web: www.acoustics.asn.au).

Preliminary Announcement

October 2006

3–6 **IEEE Ultrasonics Symposium**, Vancouver, BC, Canada (TBA).

Workshop on High Intensity Acoustic Waves in Modern Technological and Medical Applications

The First Joint Workshop of the Russian Acoustical Society (RAS) and the French Acoustical Society (SFA) on High Intensity Acoustic Waves in Modern Technological and Medical Applications was held in Moscow, Russia, on November 14 and 15, 2005. It was organized as a special session of the XVI Meeting of the Russian Acoustical Society.

The main goals of the workshop were to bring together the experts in the application of high intensity acoustic waves from France and Russia, to present the results of successful collaboration between the teams of the two countries, to initiate new contacts, and to expand the interactions between the two Societies. This workshop achieved these goals, providing wonderful opportunities for information exchange as well as for the development of long-term collaborations.

The workshop was held at Moscow State Mining University (MSU). The total number of participants was around 50. Twenty papers were presented, eleven by speakers from France and nine by speakers from Russia. The talks were grouped as follows: (1) nonlinear effects in therapeutic applications of ultrasound, (2) medical diagnostics, (3) nonlinear waves in inhomogeneous media, (4) nondestructive testing, and (5) thermoacoustics. The Proceedings and a detailed report of the workshop, as well as the English version of the Proceedings of the RAS Meeting are available online at http://ras.akin.ru

The workshop was organized by Philippe Blanc-Benson, SFA President; Vera A. Khokhlova, and Oleg A. Sapozhnikov, MSU. The Acoustical Society of America (ASA) and The International Commission for Acoustics (ICA) were two of the six sponsoring organizations.

(This report was prepared by Lawrence Crum [lac@apl.washington.edu], based on a report submitted to ICA and ASA by Vera A. Khokhlova [vera@acs366.phys.msu.ru].)

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

The purpose of these acoustical patent reviews is to provide enough information for a Journal reader to decide whether to seek more information from the patent itself. Any opinions expressed here are those of reviewers as individuals and are not legal opinions. Printed copies of United States Patents may be ordered at $3.00 each from the Commissioner of Patents and Trademarks, Washington, DC 20231. Patents are available via the Internet at http://www.uspto.gov.

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,934,218

43.30.Sf SOUNDING DEVICE FOR SHOWING ITS LOCATION ON A FISH DETECTOR

Hwang-Hsing Chen, Fort Worth, Texas
23 August 2005 (Class 367/2); filed 29 January 2004

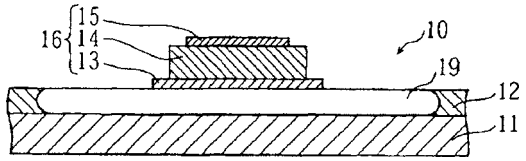
A device is described to aid a fisherman in placing the bait close to a fish or a school of fish. It comprises a receiving unit that responds to a sonar signal generated by a conventional fish finder, a wave-producing unit that reradiates a signal at the frequency of the incident sonar signal, a control unit to regulate which of these units is functioning at any given time, and a power supply unit. Presumably by placing this device near the bait so that one knows the position of the bait, it can then be moved to the position of a true fish echo.—WT

6,930,437

43.35.Ns FILM BULK ACOUSTIC RESONATOR

Tadashi Nakatani *et al.*, assignors to Fujitsu Limited
16 August 2005 (Class 310/324); filed in Japan 17 July 2001

A method of fabricating film bulk acoustic resonators is described that utilizes undercutting to acoustically isolate a resonant structure from its surroundings. One section through the device, shown in the figure, pretty much says it all. Electrodes **13** and **15** support a piezoelectric AlN layer that generates the vibrations. The conductors extend beyond the cavity **19** out of the plane of the diagram to provide contact and the cavity provides the



acoustic isolation necessary for a high-Q resonator. The conductors **13** and **15** are 100 nm thick and the AlN layer is 500 nm thick, allowing operation in the 4-GHz frequency range. Undercutting is facilitated by making the cavity in a sacrificial layer of magnesium oxide, not in the silicon substrate **11**. This, the authors argue, allows the device to be more robust than similar devices using undercutting in silicon. Little information on the electroacoustic design is given.—JAH

6,933,807

43.35.Ns ACOUSTIC REFLECTOR FOR A BAW RESONATOR PROVIDING SPECIFIED REFLECTION OF BOTH SHEAR WAVES AND LONGITUDINAL WAVES

Stephan Marksteiner *et al.*, assignors to Infineon Technologies AG
23 August 2005 (Class 333/187); filed in Germany 7 November 2002

This patent describes the use of stacks of materials having different impedances to make a perfectly reflecting barrier that isolates a bulk acoustic wave resonator from the substrate it is mounted on. It is hard to perceive anything new here, though the patent claims a receded transmissivity for longitudinal and shear waves, with beneficial results for longitudinal resonators. Nothing here seems very new—this is all based on matrix equations that have a few parameters that can be optimized on a computer of 1960s vintage.—JAH

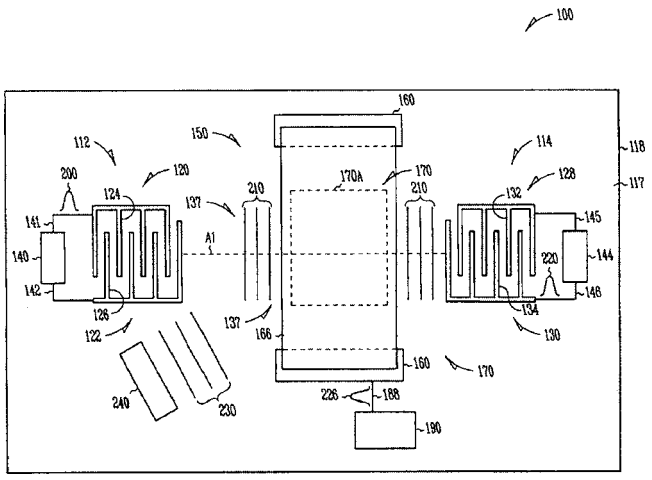
 0001-4966/2006/119(2)/681/18/$22.50

6,933,808

43.35.Ns MICROELECTROMECHANICAL APPARATUS AND METHODS FOR SURFACE ACOUSTIC WAVE SWITCHING

Qing Ma and Dong S. Shim, both of San Jose, California
23 August 2005 (Class 333/193); filed 17 July 2002

This patent describes a switchable Bragg reflector array that can be used to steer surface acoustic waves into a ''beam stop'' absorber under electrical command. The principle is quite simple: Reflector array **166** is lowered under electrostatic command and deflects lateral acoustic beam **210**



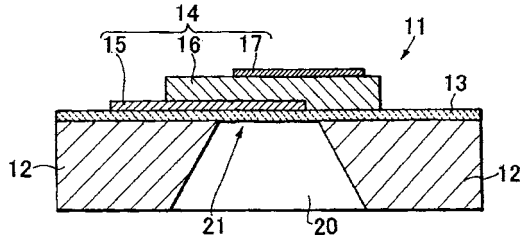
into a stop by ''electromagnetic communication.'' The effect is one of shutting off the beam, but can be used to deflect it into another circuit. The particulars are unclear, but the concepts are pretty straightforward. Will it work?—JAH

6,936,837

43.35.Ns FILM BULK ACOUSTIC RESONATOR

Tetsuo Yamada *et al.*, assignors to UBE Industries, Limited
30 August 2005 (Class 257/2); filed in Japan 11 May 2001

The authors describe a method for fabricating a thin-film resonator. This is your basic configuration of a vertically oriented plane-wave



resonator. The focus of this patent is on how to improve manufacturability of the FBAR as shown, and most of the improvements are in the materials science area.—JAH

6,940,211

43.35.Ns ELECTROACTIVE POLYMERS TRANSDUCERS AND ACTUATORS

Ronald E. Pelrine *et al.*, assignors to SRI International
6 September 2005 (Class 310/330); filed 29 June 2004

This patent describes various methods for improving the response of a piezoelectric film by prestraining or deforming it. The styles of deformation range from pleats to cylinders to flextensional types, and all have been reviewed before. Incredibly, what seems to be the authors' intent is to claim priority for all forms of prestressed polymer films. They are actually presenting a library of old ideas in polymer form.—JAH

6,930,819

43.35.Sx MINIATURIZED EXTERNAL CAVITY LASER (ECL) IMPLEMENTED WITH ACOUSTIC OPTICAL TUNABLE FILTER

Raymond R. Chu and Wilson T. C. Wong, assignors to Acceeze, Incorporated
16 August 2005 (Class 359/310); filed 3 May 2003

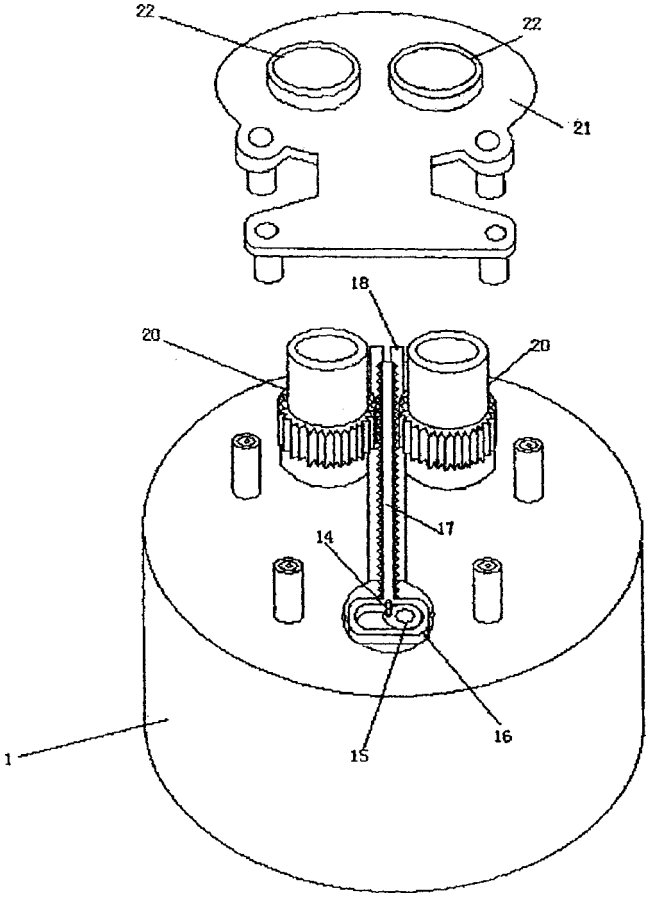
The patent relates to a new variety of external cavity laser, implemented with an acousto-optic tunable filter, to achieve high-speed, broad-range tuning with reduced channel hopping, while providing long-term reliable operations in severe environments. It is also the object of this device to provide a tunable laser with no moving parts for fiber networks, high power output, narrow laser linewidth, and high submillisecond range.—DRR

6,941,058

43.35.Sx DRIVING DEVICE FOR HANDICRAFT WITH ACOUSTO-OPTIC CONTROL AND DRIVEN BY BATTERIES

Song-Po Shih, Shenzhen, China
6 September 2005 (Class 385/147); filed 14 August 2003

This is a battery-powered device with acousto-optic control for driving a handicraft object in order to provide motion effects. A base unit contains a motor, loudspeaker, electronics, switch, and battery compartment. A cam device attached to the motor via a speed-reducing gear train engages a strip having gear teeth on each side. The strip slides in a groove, causing two



small pipes to rotate. As these pipes rotate, the light from small red-, green-, and blue-colored lamps underneath the top plate is cut off and on intermittently, producing a variety of lighting effects. Various moving displays may also be attached to the pipes. A control circuit in the base includes a shut-off timer, a sound detector circuit, and a light detector circuit.—DRR

6,931,173

43.35.Zc MEMS OPTICAL SWITCHES WITH GUARANTEED SWITCHING STATUS

Xinzhong Wang *et al.*, assignors to Alliance Fiber Optic Products, Incorporated
16 August 2005 (Class 385/18); filed 14 August 2003

Problem: Optical mirror switch elements sometimes get stuck, causing loss of data in the optical transmission path. Solution: Introduce an electrical contact into the mirror assembly that can be checked for continuity following a switching command. This confusing patent describes a system for feedback control of actuators in the optical switching arena. The text contains a lot of discussion of microprocessor control and ac versus dc impedance measurements, but there is no diagram to show what it is on the structure that is being measured, nor any analysis of the response time of such feedback and how it relates to the switching times needed for data transmission.—JAH

6,933,930

43.35.Zc ACCURATE INTERACTIVE ACOUSTIC PLATE

Fabrice Devige, Vanves and Jean-Pierre Nikolovski, Paris, both of France
23 August 2005 (Class 345/173); filed in France 29 June 2000

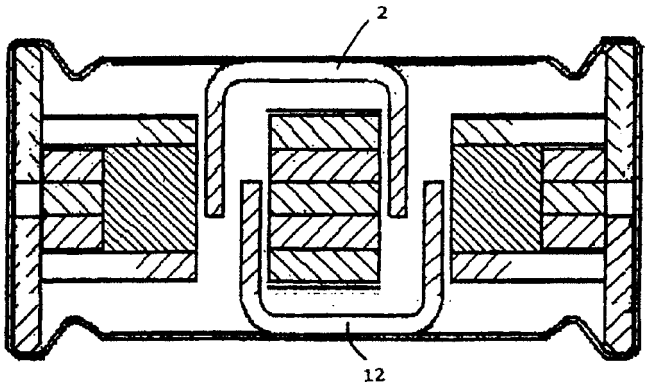
This patent describes a system for calculating the position of an impact on a piece of laminated or plate glass based upon the inputs from four piezoelectric disks situated on the corners of a rectangular plate. The method used is to time the arrivals of the symmetric-mode Lamb waves in the glass at all four transducers. There is some band limiting and peak detection processing after that, then a set of equations are solved for the x and y coordinates as a function of the time delays between pulse arrivals at the transducers located at the extremes of the x and y axes. The accuracy of the method is stated as 1 cm per 10 m, which is impressive, but this may be accomplished only by using the sound velocity as a fitting parameter.—JAH

6,931,140

43.38.Ar ELECTRO-ACOUSTIC TRANSDUCER WITH TWO DIAPHRAGMS

Aart Zeger Van Halteren and Leif Johannsen, assignors to Sonionkirk A/S
16 August 2005 (Class 381/182); filed in Denmark 11 September 2001

The patent describes the mechanical and magnetic characteristics of a dual diaphragm transducer constructed within a rectangular envelope. As the abstract states: ''When used as a microphone the transducer is substantially insensitive to vibrations, and when used as a speaker the transducer



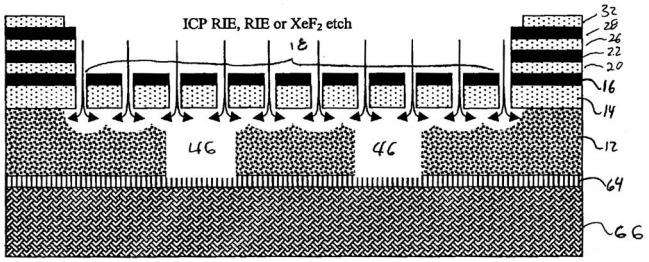
generates only very low vibration levels.'' The rectangular shape has some size advantages over conventional annular designs, and the device can also be used as a vibration generator for a silent alarm in cell phone applications.—JME

6,936,524

43.38.Ar ULTRATHIN FORM FACTOR MEMS MICROPHONES AND MICROSPEAKERS

Xu Zhu and Raymond A. Ciferno, assignors to Akustica, Incorporated
30 August 2005 (Class 438/459); filed 5 November 2003

This patent describes specific advances in micro-electro-mechanical systems (MEMS) aspects of microphone design and construction. Although



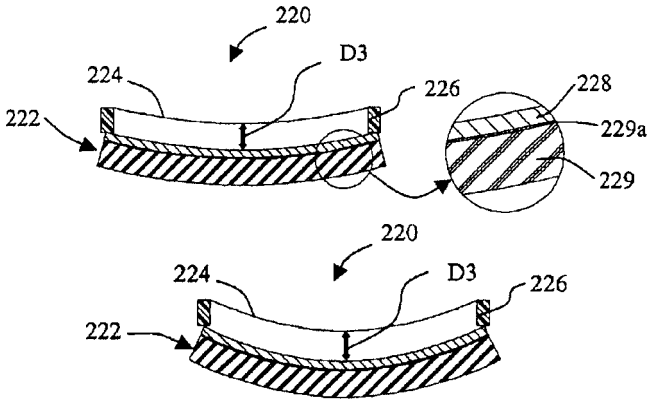
the patent discusses no direct acoustical phenomena, the scope of developments in this technology may be of general interest to many acousticians. The figure shows the complexity of what can be achieved.—JME

6,937,735

43.38.Ar MICROPHONE FOR A LISTENING DEVICE HAVING A REDUCED HUMIDITY COEFFICIENT

Dion I. de Roo *et al.*, assignors to SonionMicrotronic Néderland B.V.
30 August 2005 (Class 381/174); filed 1 August 2002

The patent describes the construction of miniature microphones for hearing aid use in which the adverse effects of humidity on microphone sensitivity are reduced by minimizing the hygroscopic coefficient of



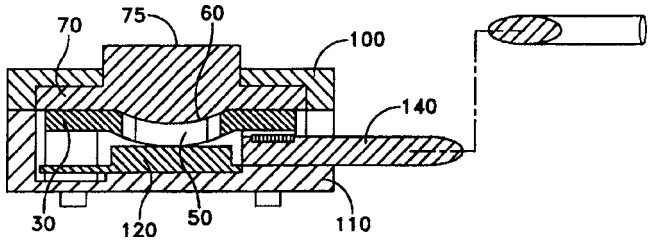
expansion in the microphone's backplate. Two bonded materials with different expansion coefficients are used to form the backplate, resulting in a reduction of expansion.—JME

6,937,736

43.38.Ar ACOUSTIC SENSOR USING CURVED PIEZOELECTRIC FILM

Minoru Toda, assignor to Measurement Specialties, Incorporated
30 August 2005 (Class 381/190); filed 5 August 2002

The patent discusses performance of contact acoustic sensors used in



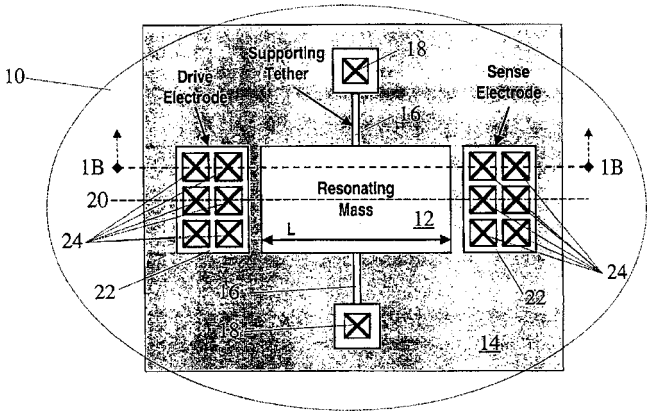
medical diagnostics. Improvements in reducing both acoustical and electrical noise are described.—JME

6,940,370

43.38.Ar MEMS RESONATOR AND METHOD OF MAKING SAME

Brian L. Bircumshaw *et al.*, assignors to The Regents of the University of California
6 September 2005 (Class 333/197); filed 6 May 2002

An interesting mechanical resonator is described that takes the form of resonant mass **12** moving in a width/length contracting-plate mode. There is



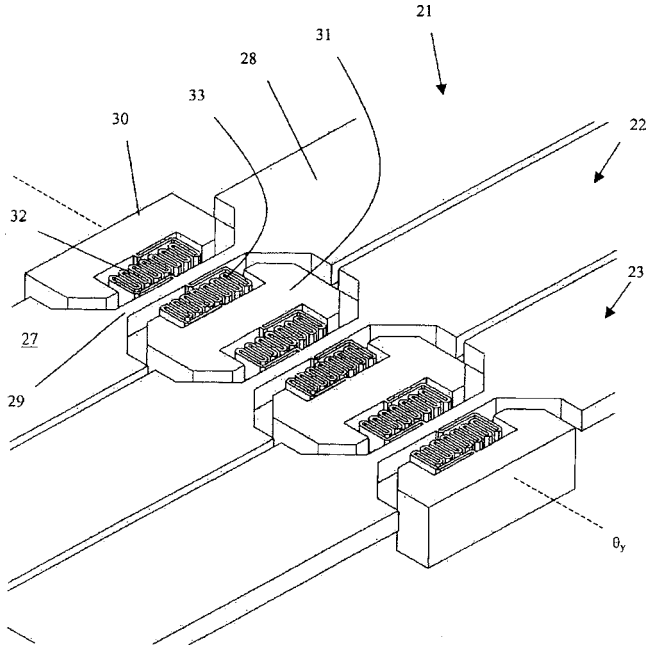
detailed discussion of an electrostatically based actuation arrangement and a fairly basic acoustic mode analysis of what is excited. The whole assembly is only microns across and resonates in the GHz frequency range.—JAH

6,934,439

43.38.Bs PIANO MEMS MICROMIRROR

Mohiuddin Mala *et al.*, assignors to JDS Uniphase Incorporated
23 August 2005 (Class 385/18); filed 27 May 2003

This patent describes an interesting simplification of a 2-D gimbal pivot for use in MEMS-optical splitters and combiners. It functions by utilizing accordion springs as the pivots as shown in the figure. Accordion pleats **33** are machined from silicon in a strictly vertical etch process to



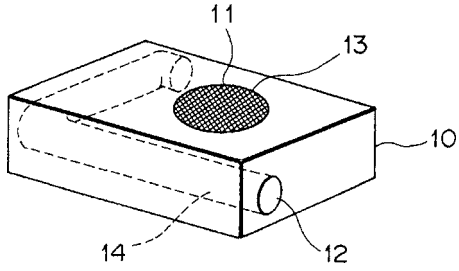
allow mirrors **21**, **22**, **23** to be independently steered in two dimensions. The authors assert that this design also allows very close packing of mirrors. This is a well written and informative patent for those interested in electrostatic manipulators.—JAH

6,912,290

43.38.Ja SPEAKER UNIT FOR LOW FREQUENCY REPRODUCTION

Erik Thorsell *et al.*, assignors to Alpine Electronics, Incorporated
28 June 2005 (Class 381/338); filed 16 November 2000

If the resonant frequency of the port assembly **12** and **14** is below the lowest frequency of interest and the resonant frequency of the loudspeaker **11** is above the highest frequency of interest, a ported enclosure results that



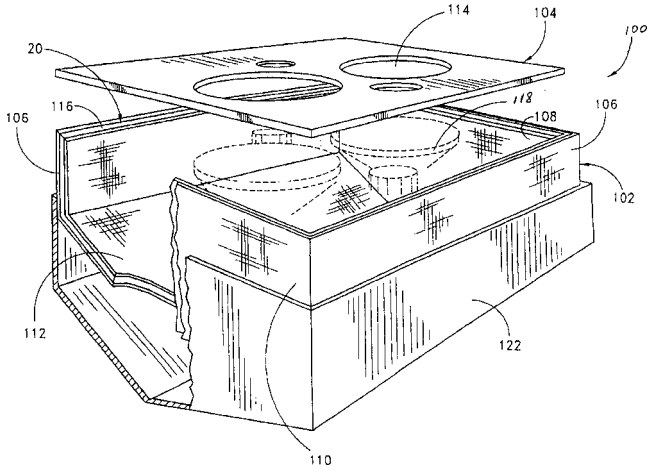
is said to exhibit ''reduced distortion and the orientation of a sound image of the lower side frequency is augmented.'' The patent is readable and interesting although it confines the discussion of the performance to the phase and magnitude of the impedance.—NAS

6,913,110

43.38.Ja LIGHTWEIGHT SPEAKER ENCLOSURE

Ross Ritto, assignor to Southern California Sound Image
5 July 2005 (Class 181/199); filed 5 August 2003

Composite loudspeaker enclosures became quite the rage in the late 1990s. The benefits of a stiff, light-weight enclosure are obvious to anyone who has lifted tour or other high-output loudspeaker enclosures made with conventional construction materials, such as plywood. The patent describes a better way to manufacture these composite enclosures and baffles than those described in some of the inventor's previous patents in this field. The



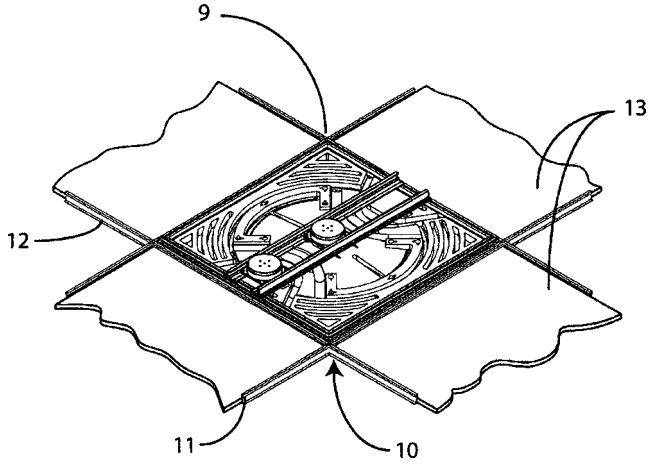
method described, vacuum bagging using a female mold, is used for many products, such as large sailboat hulls. This method is said to, and does, offer faster, and therefore a more cost efficient, means for manufacturing composite loudspeaker enclosures. The inventor is a partner in a large tour sound company.—NAS

6,929,091

43.38.Ja PLANAR DIAPHRAGM LOUDSPEAKER AND RELATED METHODS

Alejandro Bertagni *et al.*, assignors to Sound Advance Systems, Incorporated
16 August 2005 (Class 181/150); filed 27 October 2003

Long before the current interest in bending-wave loudspeakers, flat panel speakers were produced by several companies. The Bertagni name is one associated with successful commercial designs. This latest patent



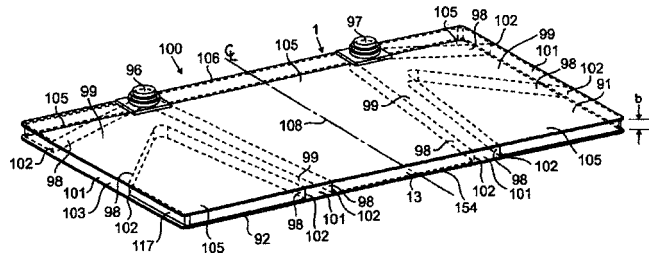
describes an improved assembly that can replace a standard lay-in ceiling tile. The goal is to mimic the appearance of adjacent tiles without compromising performance. Known techniques are combined in various ways to generate several embodiments.—GLA

6,931,143

43.38.Ja THIN ENCLOSURE ELECTROACOUSTICAL TRANSDUCING

Gerald F. Caron *et al.*, assignors to Bose Corporation
16 August 2005 (Class 381/337); filed 30 July 2002

This is a speaker system intended mainly for use in teleconferencing and can take the form of a desk pad. In a finished version, speakers **96** and **97** are enclosed behind a kind of dashboard. Front radiation is conducted through one or more channels concealed within the thickness of the pad and emerges around the perimeter of the pad. The idea is "to provide equal total



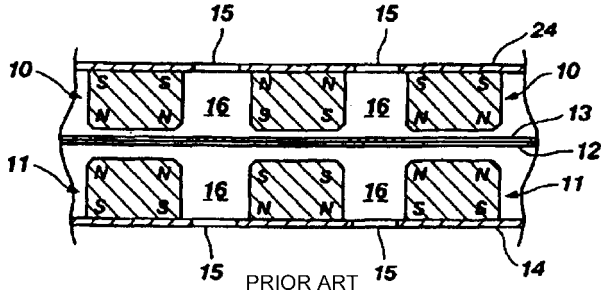
direct acoustic path lengths from an electroacoustical transducer to a predetermined point in space," thus concentrating sound at the ears of the user. (Well, it actually provides equal paths at any point along a given axis.) The idea is interesting and unusual, but at least two of the embodiments, as described in the very broad patent claims, are known prior art.—GLA

6,934,402

43.38.Ja PLANAR-MAGNETIC SPEAKERS WITH SECONDARY MAGNETIC STRUCTURE

James J. Croft III and David Graebener, assignors to American Technology Corporation
23 August 2005 (Class 381/412); filed 25 January 2002

A well-known form of planar loudspeaker utilizes a ribbon coil affixed to the diaphragm, operating in a squished-out magnetic field generated by opposing sets of magnets as shown (prior art). By making use of today's



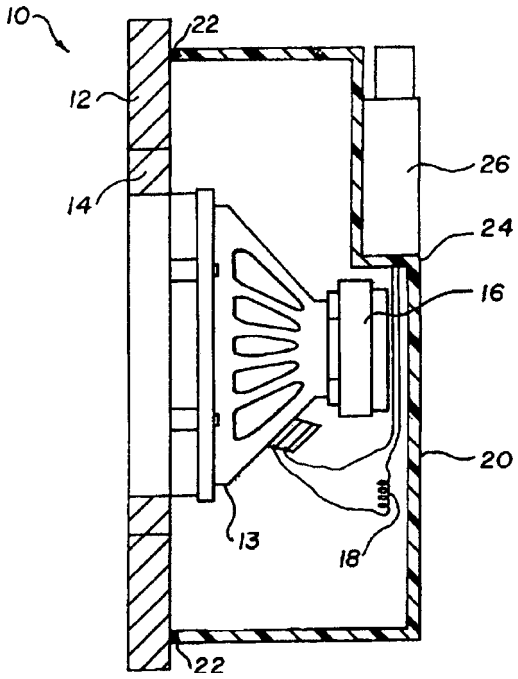
high-energy magnets it might be possible to simplify this arrangement. For example, the upper three magnets might be replaced by a single high-energy magnet. This and more than two dozen other variants are described.—GLA

6,937,740

43.38.Ja MONOPOLE LOW FREQUENCY TEST WOOFER

David Alan Dage, assignor to Visteon Global Technologies, Incorporated
30 August 2005 (Class 381/389); filed 11 April 2001

Of necessity, automotive sound systems mount loudspeakers in dashboards, door panels, and trunk lids. These all function as leaky baffles having undefined acoustical characteristics. As a result, it is difficult to design an optimum loudspeaker for such an application. This patent suggests that a



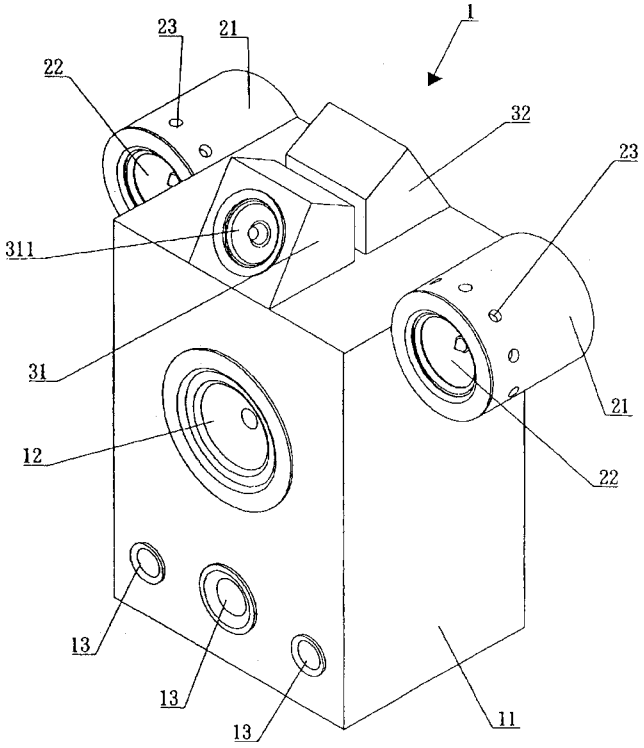
small, closed-box speaker of known performance can be temporarily installed as a test jig for comparative measurements. The idea is certainly valid. Whether it qualifies as nonobvious is another matter. In this case, the patent claims require the test box to include a recessed rear connector panel **26** and an inductor **18** in series with the woofer.—GLA

6,940,990

43.38.Ja MULTI-CHANNEL AUDIO CENTER SPEAKER DEVICE

Wan-Fang Huang, Chang Hua City, Taiwan, Province of China
6 September 2005 (Class 381/335); filed 14 July 2003

What you see here is an all-in-one L-C-R speaker system including a subwoofer. The novel feature is an additional backward-firing center speaker



32. The invention is said to ''reduce or eliminate the impact of the sound being acted directly on the audience to prevent uneasiness and broaden the range of the multi-channel audio sound.''—GLA

6,932,187

43.38.Kb PROTECTIVE ACOUSTIC COVER ASSEMBLY

Chad A. Banter and Bradley E. Reis, assignors to Gore Enterprise Holdings, Incorporated
23 August 2005 (Class 181/149); filed 14 October 2003

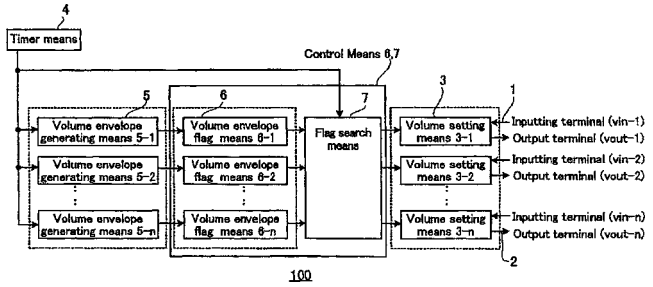
A cover that protects a microphone, loudspeaker, or other electronic acoustic transducer from moisture or oil is made from perforated metal foil whose surface is treated to make it hydrophobic or oleophobic, or both. The cover assembly described in this patent has an average maximum pore size of less than 150 μm. It has an average specific acoustic resistance of less than 11 MKS Rayls and an average specific acoustic reactance magnitude of less than about 1 MKS Rayl in the 250- to 300-Hz range, as well as an instantaneous water entry pressure value greater than about 11 cm.—EEU

6,914,990

43.38.Lc VOLUME ADJUSTING APPARATUS, METHOD AND COMPUTER PROGRAM PRODUCT

Ryoji Abe, assignor to Matsushita Electric Industrial Company, Limited
5 July 2005 (Class 381/104); filed in Japan 30 March 2001

This long and obtuse patent (written in the best patentese) apparently describes a better way of adjusting multiple channels of audio. The better



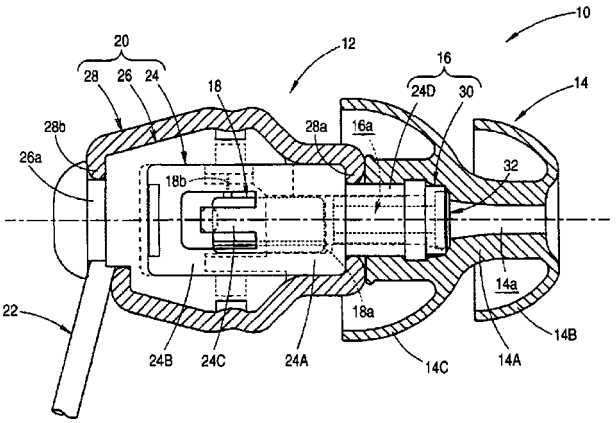
way is the addition of volume envelope flags **6** and flag search means **7** to prior art timer means **4**, volume envelope generating means **5**, and volume setting means **3**.—NAS

6,931,142

43.38.Si INSERT EARPHONE

Isao Fushimi, assignor to Star Micronics Company, Limited
16 August 2005 (Class 381/325); filed in Japan 31 October 2001

In this insert earphone, inner flange **14B** provides a comfortable air seal while outer flange **14C** supports the weight of the device. The patent's



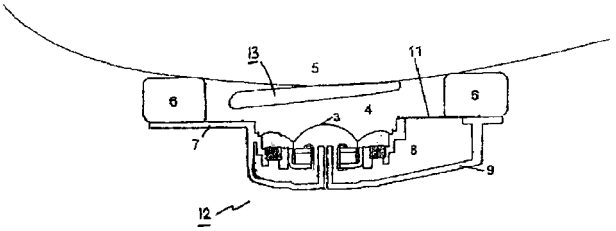
one independent claim specifies parabolic contours for both flanges, which would seem to allow plenty of leeway for competitive designs.—GLA

6,934,401

43.38.Si CLOSED HEADPHONES WITH TRANSDUCER SYSTEM

Axel Grell and Kornelia Kaddig, assignors to Sennheiser electronics GmbH & Company KG
23 August 2005 (Class 381/371); filed in Germany 24 August 2001

Contemporary closed headphones are deliberately designed with complicated air leaks to achieve optimum response during normal operation. A goal of this patent is to "simplify the damping of the fundamental resonance



of transducer systems, to improve quality and to save costs." A key factor is to make the rear volume so small that the acoustic load on the rear of the diaphragm is controlled largely by leakage resistance in the frequency range of interest.—GLA

6,937,718

43.38.Si METHOD AND APPARATUS FOR PERSONALIZED CONFERENCE AND HANDS-FREE TELEPHONY USING AUDIO BEAMING

Alexander Martin Scholte, assignor to Avaya Technology Corporation
30 August 2005 (Class 379/388.02); filed 4 September 2002

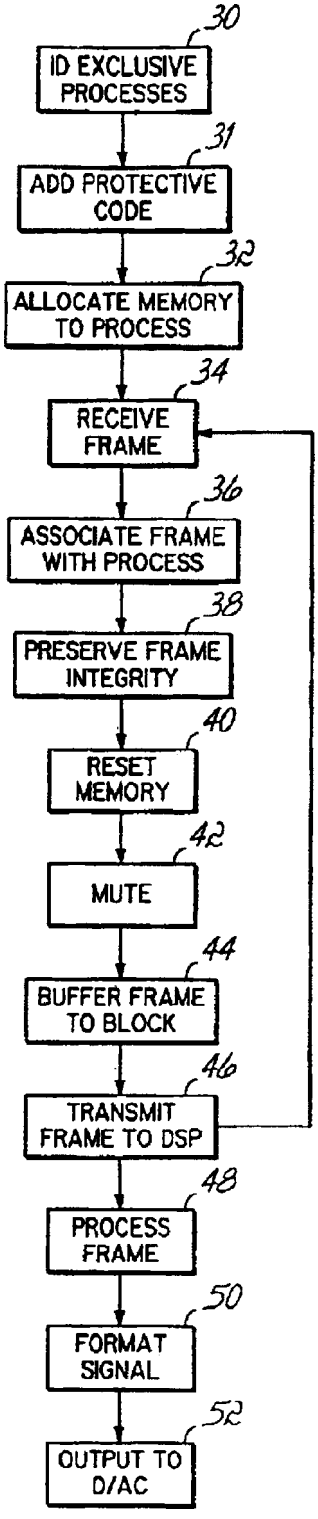
Parametric loudspeakers utilize nonlinear air transmission of ultrasonic waves to produce highly directional audible sound. Obvious applications include museum exhibits, theme park rides, point of sale exhibits, and automatic teller machines. This patent adds hands-free telephony to the list.—GLA

6,917,915

43.38.Vk MEMORY SHARING SCHEME IN AUDIO POST-PROCESSING

Robert Weixiu Du and Chinping Q. Yang, assignors to Sony Corporation
12 July 2005 (Class 704/228); filed 30 May 2001

A means is described of managing and sharing common memory for audio processes that access the data stored therein. The purpose is to



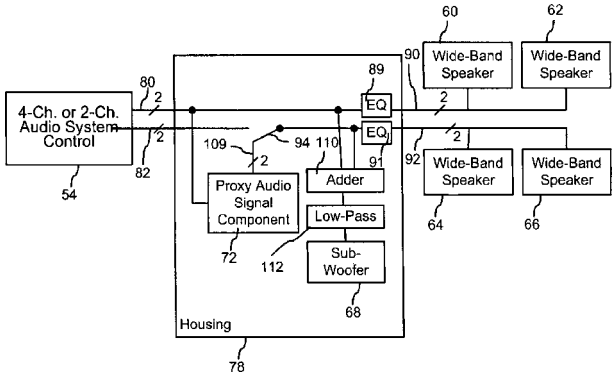
optimize the use of system resources which can reduce the cost and complexity of the memory and its management.—NAS

6,934,394

43.38.Vk UNIVERSAL FOUR-CHANNEL SURROUND SOUND SPEAKER SYSTEM FOR MULTIMEDIA COMPUTER AUDIO SUB-SYSTEMS

Jeffrey S. Anderson, assignor to Logitech Europe S.A.
23 August 2005 (Class 381/19); filed 29 February 2000

Suppose that your desktop computer setup has surround sound (four-channel) capability, including four identical loudspeakers. How does it handle a stereo (two-channel) program source? This patent teaches that an



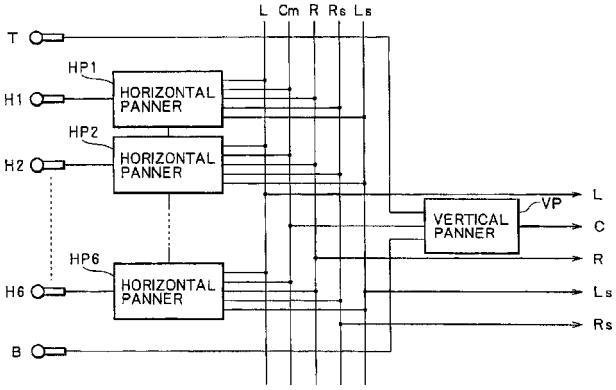
all-purpose matrix circuit generates confused imaging in this situation. An improved method is described which supports matrixed surround sound program sources, but further includes a symmetric application of left and right components to the rear speakers.—GLA

6,934,395

43.38.Vk SURROUND SOUND FIELD REPRODUCTION SYSTEM AND SURROUND SOUND FIELD REPRODUCTION METHOD

Akitaka Ito, assignor to Sony Corporation
23 August 2005 (Class 381/23); filed in Japan 15 May 2001

The patent describes a playback arrangement in which the individual signals in a three-dimensional surround sound recording can be independently panned, or assigned, to specific positions in a conventional five-



loudspeaker home surround array. The ostensible purpose is to allow the user to "fine tune" the localization of individual channels to match a variety of program input conditions.—JME

6,929,451

43.40.Tm COOLED ROTOR BLADE WITH VIBRATION DAMPING DEVICE

Shawn J. Gregg and Dominic J. Mongillo, Jr., assignors to United Technologies Corporation
16 August 2005 (Class 416/96 R); filed 19 December 2003

This rotor blade is provided with several essentially radial passages along which cooling gas flows before it is expelled along one of the blade edges. A thin rod is inserted in one of these passages, which has a series of protrusions on its inner surfaces. Friction between the rod and these protrusions, and/or rattling of the rod against these, contributes to the blade's damping.—EEU

6,935,206

43.40.Tm VIBRATION CONTROL BY CONFINEMENT OF VIBRATION ENERGY

Daryoush Alleai, assignor to Quality Research, Development & Consulting, Incorporated
30 August 2005 (Class 74/574); filed 17 October 2003

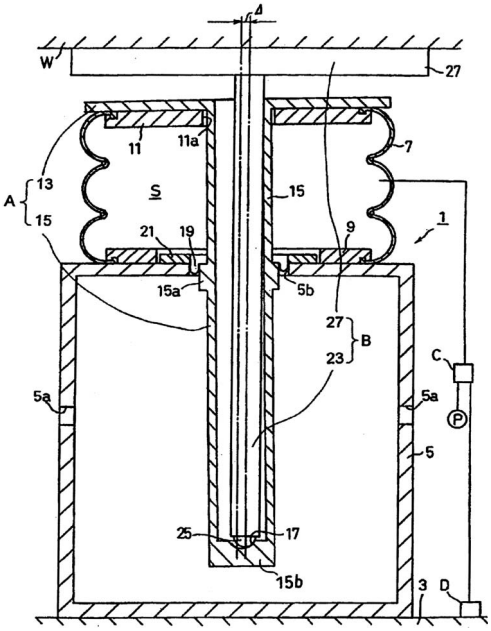
This patent is, in essence, an extension of the similarly titled United States Patent 6,871,565 [reviewed in J. Acoust. Soc. Am. **118**(3 pt 1), 1254 (2005)]. Vibrations are kept out of selected regions of a structure by means that block the transmission of vibrations from directly excited areas to these regions. Various passive and active blocking means for one-dimensional structures (beams and shafts) are delineated and applications to two-dimensional structures (plates and shells) are described.—EEU

6,935,603

43.40.Tm VIBRATION ISOLATION TABLE

Masakuni Kainuma *et al.*, assignors to Fujikura Rubber Limited
30 August 2005 (Class 248/562); filed in Japan 28 November 2002

This air-bellows isolator achieves improved isolation of horizontal vibrations by use of a base-pivoted rod **23**. The payload (typically, a sensitive instrument) rests atop a platform **27**, which is rigidly connected to the aforementioned rod. Vertical isolation is provided by a pressurized air volume **S** that is confined by elastomeric bellows **7** and by top and bottom plates. The cylindrical housing **15**, in which the base-pivoted rod **23** is free to move through a small angle, is rigidly connected to the top plate **13**, but is free to



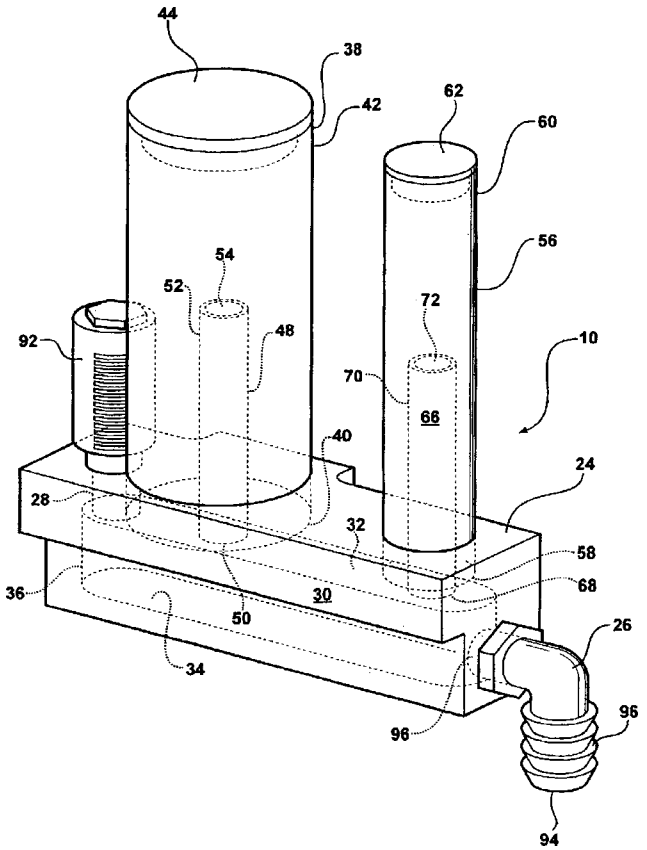
move vertically relative to the bottom plate **9**. A pneumatic seal **19** is provided between the housing and the bottom plate. The structure **5** (which has air vents **5a**) supports the assembly and enables use of a relatively long base-pivoted rod. Horizontal isolation at small excursions results from pivoting of the rod; horizontal isolation at excursions greater than Δ, at which the rod contacts the top of the cylindrical housing, is obtained by deformation of the bellows. This patent is related to patent no. 6,877,711, a review of which appeared in J. Acoust. Soc. Am. **118**(3 pt 1), 1253 (2005).—EEU

6,935,459

43.50.Gf RESONATING DEVICE FOR A PNEUMATIC SURGICAL INSTRUMENT

Tim Austin and Steve Reasoner, assignors to Stryker Instruments
30 August 2005 (Class 181/230); filed 25 February 2003

A resonating system is provided for attenuating sound waves generated by pneumatic surgical instruments. Based on the principle of the Helmholtz resonator, the system includes a manifold, at least one canister, and a neck. The manifold has a duct for accommodating a flow of fluid from a surgical instrument. Multiple canisters extend from the manifold to attenuate sound



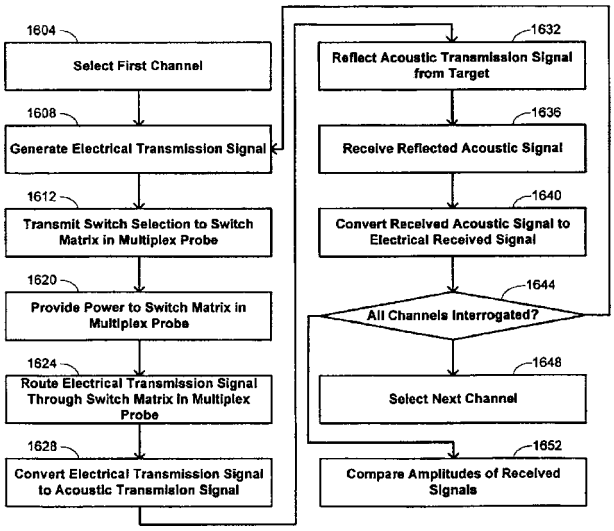
waves produced by the instrument. A neck extends into each canister and is in fluidic communication with the duct and the canister. The canisters are attached to and extend from the manifold in series, attenuating the sound waves in order from the highest frequency to the lowest frequency.—DRR

6,928,856

43.58.Vb APPARATUS AND METHODS FOR INTERFACING ACOUSTIC TESTING APPARATUS WITH ACOUSTIC PROBES AND SYSTEMS

James M. Gessert *et al.*, assignors to Sonora Medical Systems, Incorporated
16 August 2005 (Class 73/1.82); filed 5 November 2004

The patent text begins by describing how ultrasonic transducer arrays (acoustic probes) are used in medical imaging, and suggests the need for a reliable test method to verify the operation of individual transducers. The proposed method sequentially transmits an electrical test signal to each



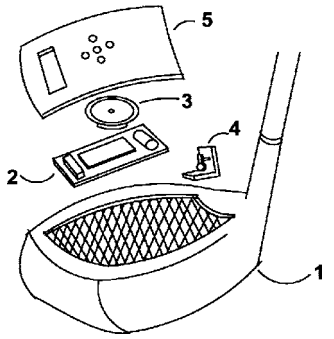
transducer, then measures the amplitude of the return signal or analyzes it in some other way. As just described, the process seems obvious. However, the patent claims require that the test signal be created and routed through a narrowly defined arrangement of adapter channels, switch-matrix channels, and relay elements.—GLA

6,935,965

43.58.Wc GOLF CLUB DEVICE AND APPARATUS FOR TEACHING GOLF SWING RHYTHM AND TEMPO

Craig Neil DeVarney, Essex Junction, Vermont
30 August 2005 (Class 473/224); filed 19 April 2004

The duffers are at it again. This time the idea is to include the ubiquitous sound chip inside the club and play a sound at the beginning of the



swing. This might possibly help aspects of timing but issues of stance and posture and ball contact are not touched by this invention. What would Jack Nicklaus think?—MK

6,934,364

43.60.Bf HANDSET IDENTIFIER USING SUPPORT VECTOR MACHINES

Purdy PinPin Ho, assignor to Hewlett-Packard Development Company, L.P.
23 August 2005 (Class 379/21); filed 28 February 2002

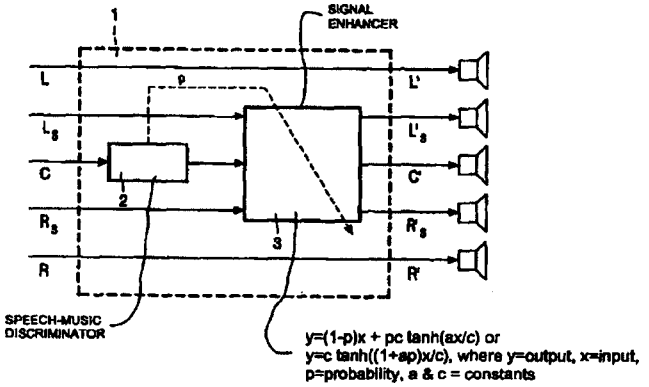
General signal classification using support vector machines (SVMs) is employed here to classify the signal source transmitter, or "handset," from which a speech signal originates. The number of SVMs required matches the number of possible handsets, since each SVM typically yields a binary decision over a multidimensional feature space. It is suggested that the usual mel-frequency cepstral vectors can be used here, together with commercially available SVM training software. The general idea of using multiple SVMs for multiple signals seems to be the true object of the patent.—SAF

6,914,988

43.60.Dh AUDIO REPRODUCING DEVICE

Roy Irwan and Erik Larsen, assignors to Koninklijke Philips Electronics N.V.
5 July 2005 (Class 381/22); filed in the European Patent Office 6 September 2001

This short patent with one figure describes a means of processing multichannel audio data to enhance the speech portion of the signal in the center and surround channels of a multichannel stream. The technique uses the nonlinear discriminator described in "A Real Time Speech Music Discriminator" by Aarts and Dekker in J. Audio Eng. Soc. **47**(9), 720–725



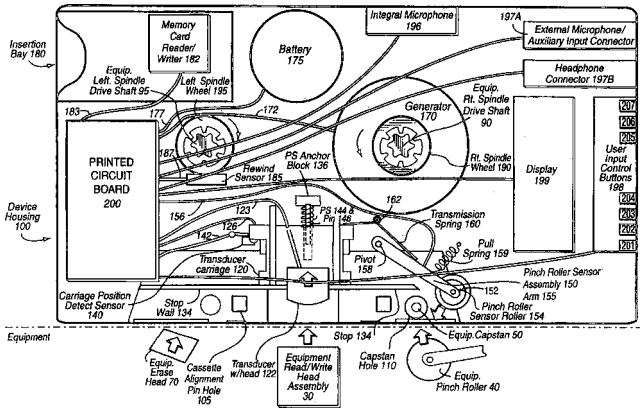
(2005), which controls the enhancing scheme described in WO 02/50831 A2. By means of a probability, p, determined by the speech-music discriminator, the signal enhancer acts on the center and surround channels or the center channel alone using hyperbolic tangent functions. The claims portion in the patent takes up about one-third of the five total written columns.—NAS

6,941,180

43.60.Dh AUDIO CASSETTE EMULATOR

Addison M. Fischer and Robert L. Protheroa, both of Naples, Florida
6 September 2005 (Class 700/94); filed 29 July 1999

This rather interesting device, having precisely the same shape as a standard audio cassette, accepts digital audio data from a variety of sources, e.g., Internet transmission, a digital computer, or flash memory card, and plays the digital data through any standard audio cassette player. The user can control the playback using the equipment's standard controls, START, STOP, PLAY, REWIND, FAST FORWARD, etc. An example of its use is the playback on analog equipment of music or other type of performance recorded in digital format. The device functions by converting the digital representation of the sound into magnetic signals that are coupled into the read/write head of the cassette player. The device includes at least one digital processor and a slot into which a memory card may be inserted.—DRR

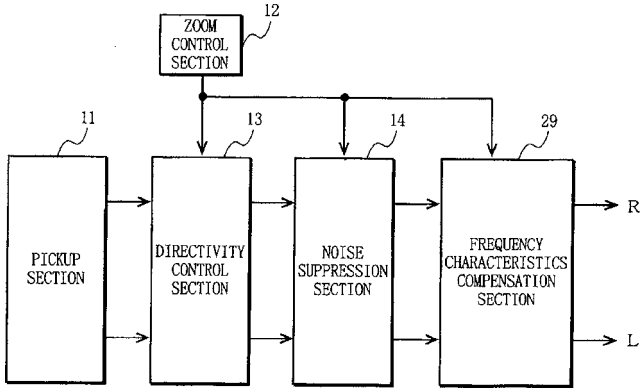


6,931,138

43.60.Fg ZOOM MICROPHONE DEVICE

Takashi Kawamura *et al.*, assignors to Matsushita Electric Industrial Company, Limited
16 August 2005 (Class 381/92); filed in Japan 25 October 2000

This interesting and well-written patent describes a "zoom" microphone array suitable for use in conjunction with a zoom video camera. The input section contains a number of basic microphone elements and provides for synthesis of patterns from wide cardioid (subcardioid) with a DI of 2.5 dB, up to second-order cardioid with DI values in the range of 9 to 10 dB.



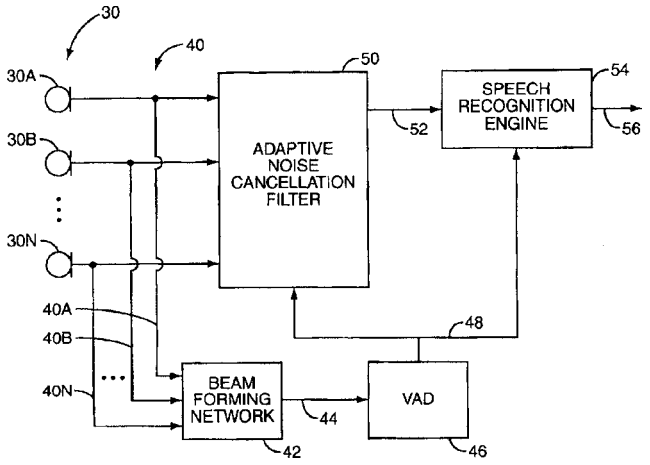
As the camera zoom function varies from wide angle to narrow angle, the microphone pattern will range from low DI to high. There will of course be a reduction in level pickup, and noise will increase as these functions are carried out. The system further equalizes level and spectrum to match the picture, and the noise floor is reduced accordingly.—JME

6,937,980

43.60.Fg SPEECH RECOGNITION USING MICROPHONE ANTENNA ARRAY

Leonid Krasny *et al.*, assignors to Telefonaktiebolaget LM Ericsson (publ)
30 August 2005 (Class 704/231); filed 2 October 2001

The patent describes a combination of useful techniques in hands-free automotive communications. Multimicrophone beamforming, operating in



conjunction with an adaptive noise-reducing system, provides input to a ''speech recognition engine'' that further enhances intelligibility.—JME

6,935,335

43.64.Ri SYSTEM AND METHOD FOR TREATING OBSTRUCTIVE SLEEP APNEA

Michael L. Lehrman and Michael E. Halleck, assignors to iLife Systems, Incorporated
30 August 2005 (Class 128/200.24); filed 17 August 2000

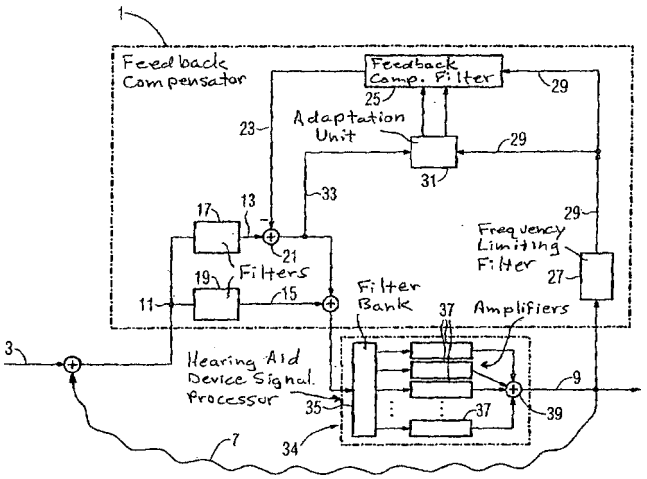
This device consists of one or more microphones, placed in the vicinity of the neck, to detect breathing sounds within the airway of a sleeping person. The microphone signals are relayed to a controller, which applies digital signal processing to identify at least one breathing pattern that occurs at the onset of an obstructive sleep apnea event. When the controller detects such an event, it sends an alarm signal to a stimulus generator. The stimulus generator creates a stimulus (an electric current, a sound, a vibrator, a flashing light, etc.) that causes the sleeping person to move in a manner to terminate the obstructive sleep apnea event before the cessation of breathing occurs. The patent asserts that this device can terminate a sleep apnea event without causing psychological stress associated with the cessation of breathing.—DRR

6,931,137

43.66.Ts FEEDBACK COMPENSATION DEVICE AND METHOD, AND HEARING AID DEVICE EMPLOYING SAME

Georg-Erwin Arndt and Tom Weidner, assignors to Siemens Audiologische Technik GmbH
16 August 2005 (Class 381/83); filed in Germany 30 September 2002

An acoustic feedback reduction scheme is described in which the input signal is split into two paths to reduce signal processing complexity and the



amount of buffer storage required. The system may be utilized for multichannel acoustic devices.—DAP

6,931,141

43.66.Ts HEARING AID AND A METHOD FOR OPERATING A HEARING AID

Finn Thoger Moller, assignor to GN Resound A/S
16 August 2005 (Class 381/323); filed in Denmark 12 October 2001

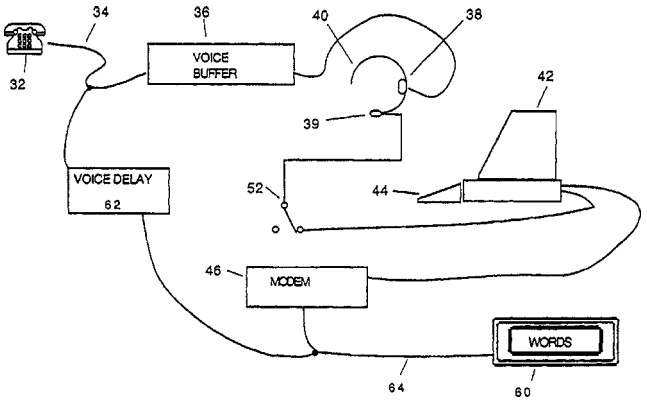
Methodology is provided for a hearing aid to detect when an external device such as a programmer or audio source is connected or disconnected. When such a device is connected, the hearing aid may be brought into a mode to receive programming or audio data. When disconnected, the hearing aid returns to normal processing mode. When unconnected, the closely spaced connector terminals are kept at nearly the same electrical potential to prevent corrosion.—DAP

6,934,366

43.66.Ts RELAY FOR PERSONAL INTERPRETER

Robert M. Engelke *et al.*, assignors to Ultratec, Incorporated
23 August 2005 (Class 379/52); filed 10 December 2002

This is a relay system for telephone communication between hearing-impaired users and normal-hearing users with a call assistant as a go-between. Its purpose is to minimize typing by the call assistant of material to be sent as text to the hearing-impaired user. Instead, the call assistant



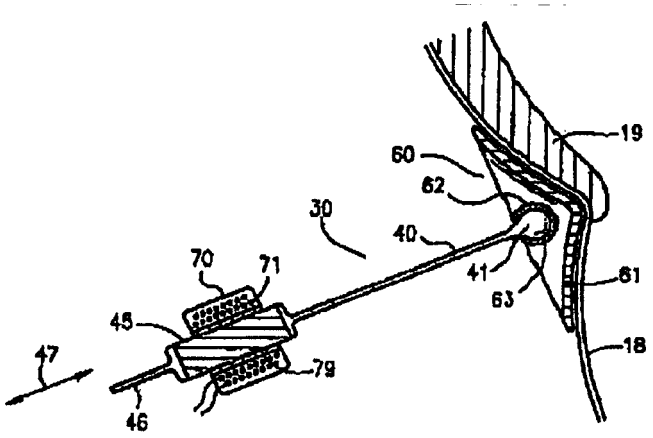
revoices the words spoken by the hearing user into a voice-recognition system trained to the voice of that call assistant. The text stream generated by the computer and the voice of the hearing user are both sent to the assisted user, who is now supplied with a visual text stream to supplement the voice communication.—DRR

6,934,400

43.66.Ts METHOD FOR CONTROLLING THE DYNAMIC RANGE OF A HEARING AID, AND METHOD TO MANUFACTURE DIFFERENT HEARING AIDS, AND A HEARING AID

Andi Vonlanthen, assignor to Phonak AG
23 August 2005 (Class 381/320); filed 3 November 2000

The dynamic range of a hearing aid, cochlear implant, or earphone is controlled manually or automatically by switching the input impedance of the acoustic-mechanical output transducer in accordance with the wearer's hearing characteristics and/or the acoustic environment.—DAP

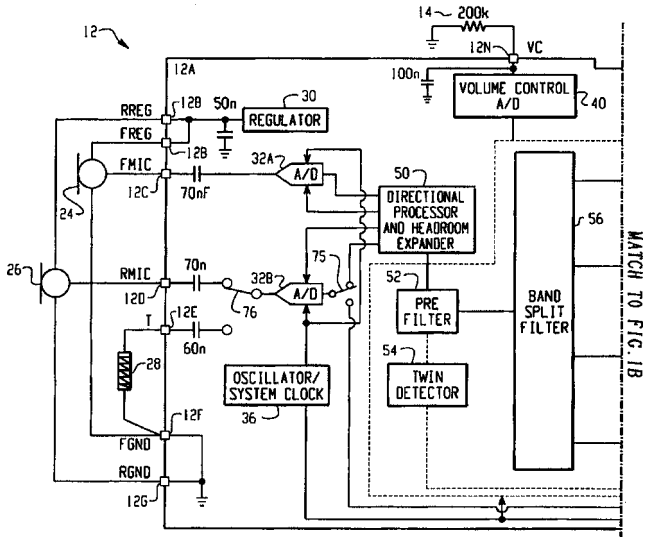


6,937,738

43.66.Ts DIGITAL HEARING AID SYSTEM

Stephen W. Armstrong *et al.*, assignors to Gennum Corporation
30 August 2005 (Class 381/312); filed 12 April 2002

In a multimicrophone, directional, digital hearing aid, occlusion is reduced by subtracting the rear-microphone-channel signal pickup of the unwanted wearer's own voice in the ear canal from the processed, intended signal. Headroom expansion is accomplished by dynamically adjusting the gain before and after the analog-to-digital converters.—DAP



6,940,989

43.66.Ts DIRECT TYMPANIC DRIVE VIA A FLOATING FILAMENT ASSEMBLY

Adnan Shennib *et al.*, assignors to InSound Medical, Incorporated
6 September 2005 (Class 381/326); filed 30 December 1999

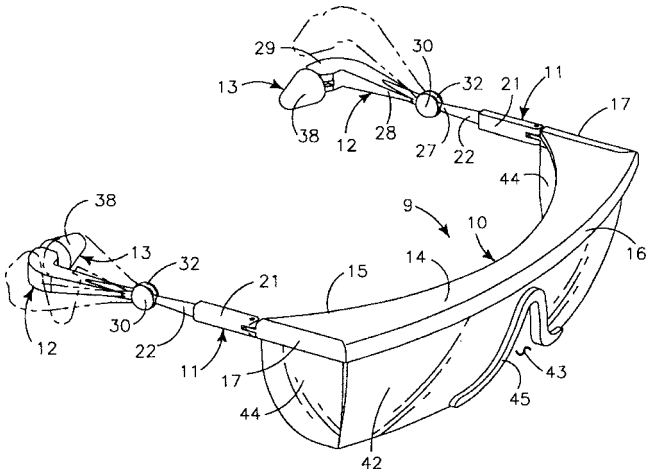
A nonoccluding hearing aid positioned deep in the ear canal stimulates the tympanic membrane directly via a miniature, floating, elongated, magnetic, vibrational filament. Transduction occurs via magnetic-to-vibrational energy conversion by a coil wrapped around the magnetic filament. Stimulation is said to be consistent regardless of the exact position of the hearing device within the ear canal.—DAP

6,728,974

43.66.Vt SAFETY GOGGLES WITH EARPLUGS

Jake Wadsworth, St. Ignatius, Montana
4 May 2004 (Class 2/456); filed 8 October 2002

Earplugs are mounted on the ends of temple pieces of safety glasses. The temples are fitted with a universal joint to permit multiple degrees of freedom for adjustment of the positioning of the earplugs.—JE

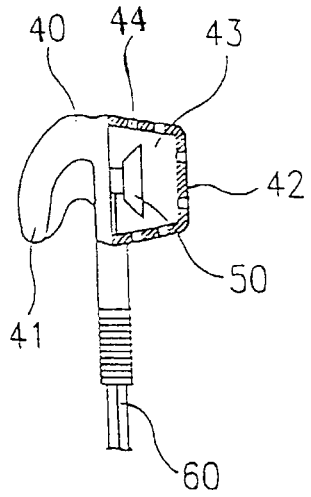


6,751,328

43.66.Vt EARPHONE WITHOUT IMPULSE NOISE FOR PROTECTION AGAINST CONDUCTIVE HEARING LOSS

Chung-Yu Lin, Kaohsiung, Taiwan, Province of China
15 June 2004 (Class 381/330); filed in Taiwan, Province of China 13 January 1998

Based on an unusual theory of the level and type of hearing pathology caused by impulsive noise, this device purports to eliminate such noise while preserving the quality of sound. This is accomplished by hanging an



enclosed loudspeaker **40** from an earplug **41** and directing the sound outwards to be partially absorbed by the external ear structures. This is said to prevent conductive hearing loss. I guess sensorineural hearing losses are never a problem.—JE

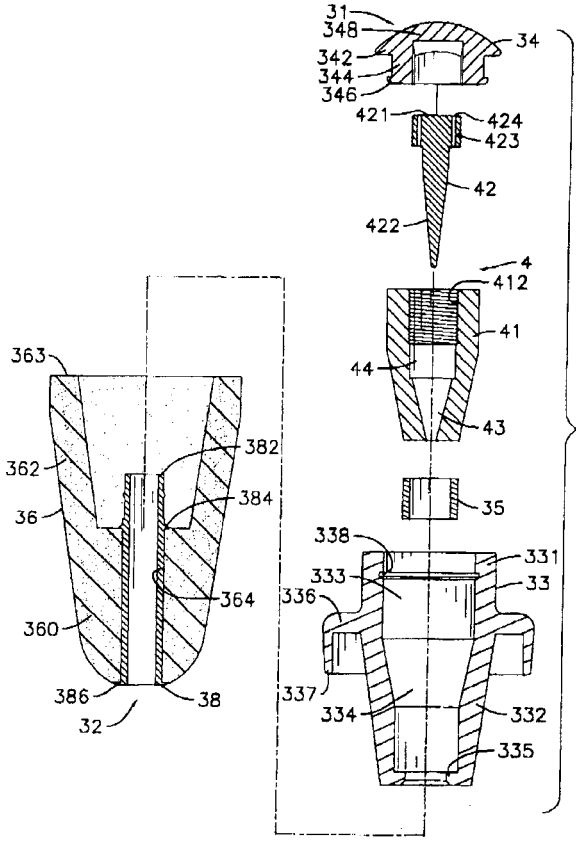
6,938,622

43.66.Vt EARPLUG AND APPARATUS WITH THE EARPLUGS

Fu-Sheng Huang, Chungho City, Taipei Hsien, Taiwan, Province of China

6 September 2005 (Class 128/864); filed in Taiwan, Province of China 24 October 2003

This complex hearing protector consists of six parts including a replaceable soft part **36** and an insert **42** that can be adjusted to provide different degrees of attenuation. The cap **34** can be replaced by a transducer



for communication. The objectives of the invention are to provide an instantly available, mass producible earplug that can be custom fit by "just applying some replaceable elements."—JE

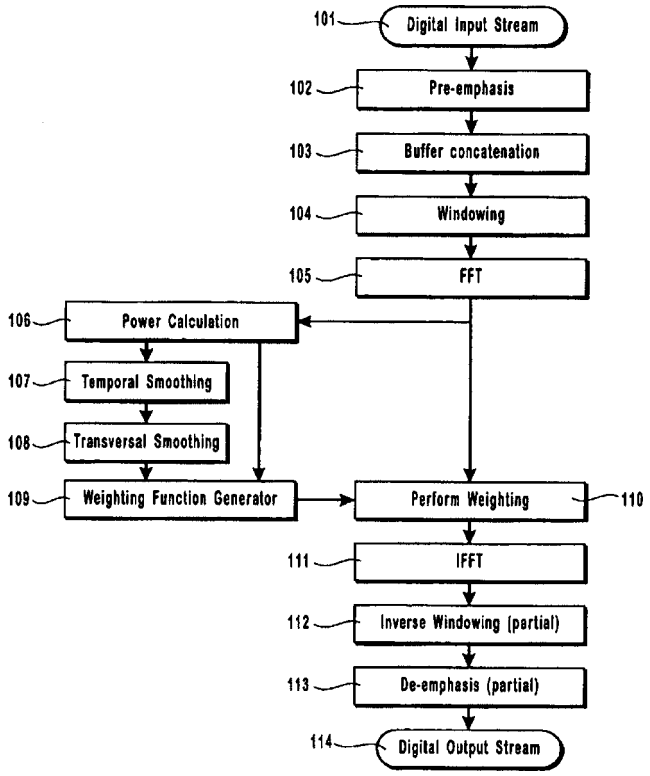
6,931,292

43.72.Dv NOISE REDUCTION METHOD AND APPARATUS

Marcia R. Brumitt and James M. Turnbull, assignors to Jabra Corporation

16 August 2005 (Class 700/94); filed 19 June 2000

This patent describes a method for spectrum-domain background noise reduction of a communication signal such as speech. Refering to the figure, one can see that a simple transform—noise reduction—inverse transform procedure is envisioned, wherein the spectrum weighting functions are



crucial. "The purpose of the weighting function is to leave the frequency bins with relatively large power levels unchanged and to attenuate the frequency bins with relatively low power levels." This approach is as naive as it is simple, and will obviously work only for noise with a lower level than the important speech components.—SAF

6,934,681

43.72.Fx SPEAKER'S VOICE RECOGNITION SYSTEM, METHOD AND RECORDING MEDIUM USING TWO DIMENSIONAL FREQUENCY EXPANSION COEFFICIENTS

Tadashi Emori and Koichi Shinoda, assignors to NEC Corporation

23 August 2005 (Class 704/250); filed in Japan 26 October 1999

A method is described which addresses the speaker normalization problems affecting traditional speaker-independent speech recognition algorithms. Quite often, HMMs for speech units are trained over Gaussian mixture models using many speakers, leading to a sort of monolithic model of every speaker's sounds. These models may invoke frequency-domain warping using the best selection of preset parameters, in an effort to better match

a particular speaker. This patent describes, in sufficient detail, several techniques for incorporating cepstrum-domain, frequency-spectrum warping to provide rudimentary speaker normalization as an additional trainable parameter in the HMMs, thereby customizing the process for each training speaker and avoiding the selection from preset warping parameters.—SAF

6,934,370

43.72.Gy SYSTEM AND METHOD FOR COMMUNICATING AUDIO DATA SIGNALS VIA AN AUDIO COMMUNICATIONS MEDIUM

Roy Leban *et al.*, assignors to Microsoft Corporation
23 August 2005 (Class 379/102.01); filed 16 June 2003

A computer receives voice and audio data signals from a telephone speaker via its microphone, extracts an event message, and performs an action based on the event message. The computer generates and plays an audio data signal corresponding to the action it performed.—DAP

6,931,255

43.72.Ja MOBILE TERMINAL WITH A TEXT-TO-SPEECH CONVERTER

Fisseha Mekuria, assignor to Telefonaktiebolaget L M Ericsson (publ)
16 August 2005 (Class 455/466); filed 19 September 2001

The patent presents the idea of employing a ''low-complexity'' text-to-speech algorithm to provide speech synthesis of text messages on a low-capacity mobile device. It is suggested that a ''scaled version of a rule-based, text-to-speech synthesis system without the requirement of speech naturalness'' can be used. No further details of the TTS method are provided, suggesting that no such system has actually been tried, yet the missing TTS method is advertised to provide ''highly intelligible voice signals.''—SAF

6,941,269

43.72.Ja METHOD AND SYSTEM FOR PROVIDING AUTOMATED AUDIBLE BACKCHANNEL RESPONSES

Harvey S. Cohen and Kenneth H. Rosen, assignors to AT&T Corporation
6 September 2005 (Class 704/275); filed 23 February 2001

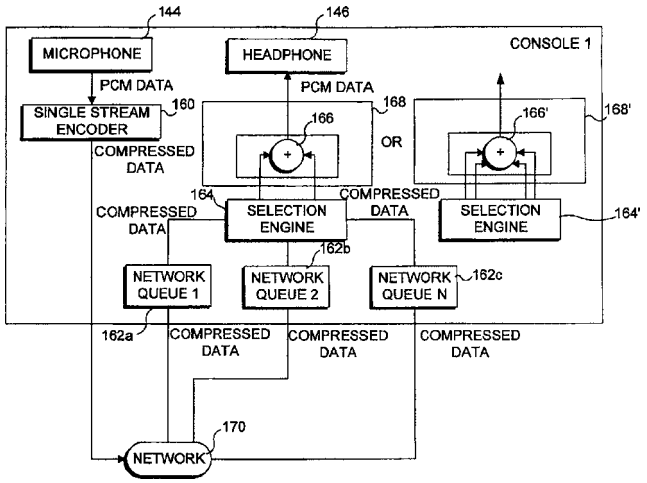
This patent addresses a problem of human behavior, namely that someone talking to a machine quickly tires of doing so, partly because they receive no backchannel communication, such as ''uh huh,'' ''gotcha,'' etc. The methods here involve identifying pauses in user-input speech and filling those pauses with predetermined, synthesized, backchannel items such as ''got it'' and ''okie dokie [sic].'' One can only guess at the potential reaction to an automated device saying ''okie dokie.''—SAF

6,935,959

43.72.Kb USE OF MULTIPLE PLAYER REAL-TIME VOICE COMMUNICATIONS ON A GAMING DEVICE

Damon V. Danieli and Roxana Gabriela Arama, assignors to Microsoft Corporation
30 August 2005 (Class 463/42); filed 16 May 2002

According to the inventors, verbal reparteé is missing from today's video games. Accordingly, the disclosure lays out a networked scheme using



peer-to-peer communication. As seen from the figure, speech is compressed and transmitted to other players. Their packets are selected and presented to the user.—MK

6,941,161

43.72.Kb MICROPHONE POSITION AND SPEECH LEVEL SENSOR

James F. Bobisuthi *et al.*, assignors to Plantronics, Incorporated
6 September 2005 (Class 455/569.1); filed 13 September 2001

A procedure is described for detecting when the volume of a speaker's voice at a microphone is insufficient for good transmission, either because it provides insufficient signal-noise ratio over background, or because it is simply too quiet for the microphone. The speaker is then warned of the problem by a visual indicator, and is thus given the opportunity to make the appropriate adjustments to their headset mic or whatever is in use. Methods are described for waiting an appropriate interval between loud speech amplitude peaks before sending the warning.—SAF

6,928,404

43.72.Ne SYSTEM AND METHODS FOR ACOUSTIC AND LANGUAGE MODELING FOR AUTOMATIC SPEECH RECOGNITION WITH LARGE VOCABULARIES

Ponani Gopalakrishnan *et al.*, assignors to International Business Machines Corporation
9 August 2005 (Class 704/10); filed 17 March 1999

In this speech recognition system, language modeling is done using a combination of words and subword morphemes. Intended especially for highly inflected languages, such as Russian, the method allows the use of n-gram frequency statistics in cases that would be prohibitive using only word modeling. The decision of whether a word is to be modeled as a whole or by morphemes is based primarily on word-frequency statistics. The first claim sets out a sequence of steps to be used in splitting up a word.—DLR

6,928,614

43.72.Ne MOBILE OFFICE WITH SPEECH RECOGNITION

Charles Allen Everhart, assignor to Visteon Global Technologies, Incorporated
9 August 2005 (Class 715/728); filed 13 October 1998

This automotive speech recognition system is intended to find use as an ''office on wheels,'' with emphasis on the design of the dialogue interaction to simplify the use of email, browser, and various information-access

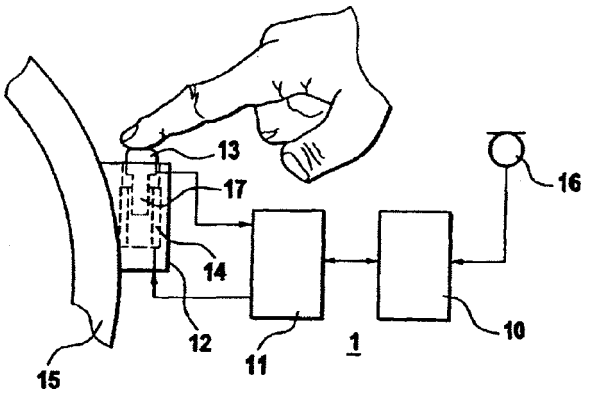
functions, all while watching for red lights and traffic cops—not to mention other cars. Commercially available synthesis and recognition packages are assumed. The interface design includes a set of steering-wheel buttons, which allow instant system customization for experienced users.—DLR

6,930,265

43.72.Ne METHOD AND SYSTEM FOR ACOUSTIC FUNCTION CONTROL IN MOTOR VEHICLES

Winfried Koenig, assignor to Robert Bosch GmbH
16 August 2005 (Class 200/308); filed in Germany 29 July 2000

The concept discussed here proceeds from several *a priori* assumptions, any of which may be invalid. First is that a speech recognition system in an automobile will require a ''push-to-talk'' signal. Second is that the user



will require feedback that such a signal has been processed by the system. And third is that the user, while driving a car, may not notice that feedback. Accordingly, the disclosed device is a special button connected so as to vibrate as a form of feedback.—DLR

6,931,374

43.72.Ne METHOD OF SPEECH RECOGNITION USING VARIATIONAL INFERENCE WITH SWITCHING STATE SPACE MODELS

Hagai Attias *et al.*, assignors to Microsoft Corporation
16 August 2005 (Class 704/240); filed 1 April 2003

It is rare to see a truly novel approach to speech recognition these days, so this patent is exceptional on that basis alone. ''Speech is represented as the output of an attempt by the speaker to phonetically implement a linguistic definition of a sequence of speech units,'' which are in turn represented (e.g., by vocal tract resonances) within a dynamic trajectory model implemented as a state-space formulation with recursive noise added at each frame. The trajectory itself is hidden, and so is a hidden trajectory model. Unfortunately, the EM algorithm for such a model is not tractable, and so major innovations here include methods for approximating its posterior using either a Gaussian mixture or an HMM posterior for training. Given training on one of these, the hidden production-related parameters which result can be simply recovered.—SAF

6,934,682

43.72.Ne PROCESSING SPEECH RECOGNITION ERRORS IN AN EMBEDDED SPEECH RECOGNITION SYSTEM

Steven G. Woodward, assignor to International Business Machines Corporation
23 August 2005 (Class 704/250); filed 1 March 2001

Here is yet another patent which addresses a popular problem of late, viz, what to do with misrecognized speech. While a number of recent patents have invoked the text domain for presenting the user with several options to select for the problematic word (as is also patented here), a further embodiment would engage the user in a dialogue so he/she may be presented with the recognition options in the speech modality. This reviewer is beginning to wonder how many times this particular wheel can be invented.—SAF

6,934,685

43.72.Ne VOICE RECOGNITION DEVICE FOR TOYS

Takashi Ichikawa, assignor to Toytec Corporation
23 August 2005 (Class 704/275); filed in Japan 21 April 1999

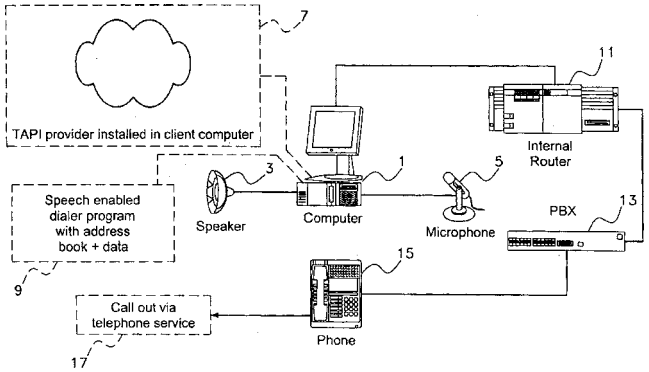
A very simple approach is taken to permit speaker-independent recognition of a (presumably) limited vocabulary by a toy. The processor uses basic pattern classification just on the varying lengths of speech signals. This leads to extremely low computational complexity in comparison to actual speech recognition schemes, but the supposed result that ''the speech of any person can be recognized irrespective of sex, age, or the like'' seems optimistic and is not substantiated in the patent.—SAF

6,940,951

43.72.Ne TELEPHONE APPLICATION PROGRAMMING INTERFACE-BASED, SPEECH ENABLED AUTOMATIC TELEPHONE DIALER USING NAMES

Jerome R. Mahoney, assignor to iVoice, Incorporated
6 September 2005 (Class 379/88.03); filed 23 January 2002

A speech-enabled automatic name dialer for connection to a telephone system utilizes a computerized address book to store name/telephone-



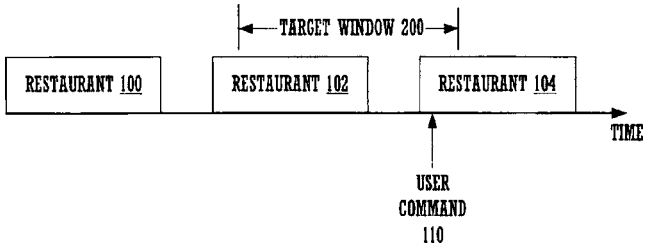
number data. Software creates phonemes from names of data sets and initiates dialing after matching the voice input phonemes to the stored data.—DAP

6,941,268

43.72.Ne HANDLING OF SPEECH RECOGNITION IN A DECLARATIVE MARKUP LANGUAGE

Brandon W. Porter *et al.*, assignors to TellMe Networks, Incorporated
6 September 2005 (Class 704/270); filed 21 June 2001

A speech recognition system used for list navigation determines the time at which a user speaks before queued audio prompts are finished playing. No *a priori* knowledge about the length of each prompt is required.



Markers are placed into the program code which allows an overlapping target window to be created that extends into the start of playback of subsequent items. Decisions are made by comparing time from last mark to time from last command.—DAP

6,931,377

43.75.Rs INFORMATION PROCESSING APPARATUS AND METHOD FOR GENERATING DERIVATIVE INFORMATION FROM VOCAL-CONTAINING MUSICAL INFORMATION

Kenji Seya, assignor to Sony Corporation
16 August 2005 (Class 704/277); filed in Japan 29 August 1997

It's not enough to have karaoke. The issue is what is the language of the singer? What this inventor proposes is a scheme for removal of the voice from the musical track, word recognition and translation to another language, and, finally, the resynthesis of the singing voice combined with the instrumental track. Since all the signal processing details are absent, and since all of these are very hard problems, one can only assume the patent is conceptual.—MK

6,930,236

43.75.Wx APPARATUS FOR ANALYZING MUSIC USING SOUNDS OF INSTRUMENTS

Doill Jung, assignor to Amusetec Company, Limited
16 August 2005 (Class 84/616); filed in the Republic of Korea 18 December 2001

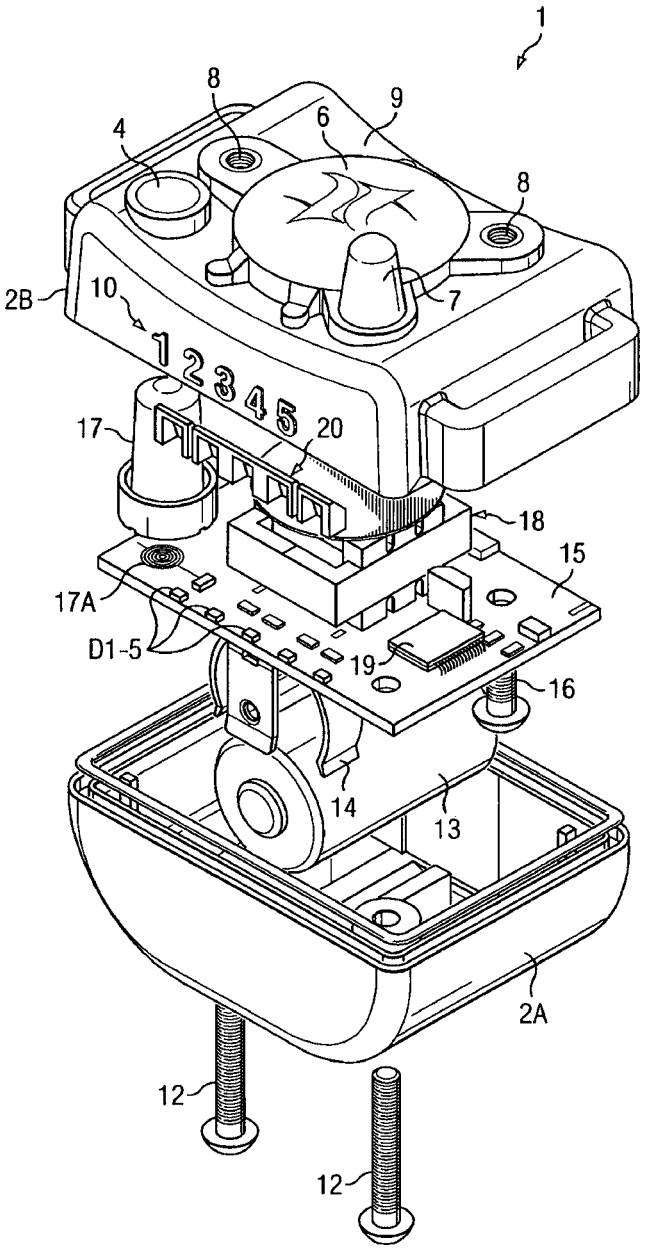
Fundamentally, this discloses a method for doing spectral subtraction of monophonic musical instruments from a possibly multiphonic source. There are several complexities which the text glosses over: How is the best match between the input and library to be found? How can this method function when the overtone series forms new timbres (*a la* Ravel)? For this method to work, the source material must have nicely separated spectral peaks and this will be difficult to obtain in practice for real multiphonic sources.—MK

6,928,958

43.80.Ka VIBRATION SENSOR ASSEMBLY AND METHOD FOR BARK CONTROLLER

Timothy J. Crist *et al.*, assignors to Tri-Tronics, Incorporated
16 August 2005 (Class 119/718); filed 7 January 2004

This device is a collar-mounted electronic ''bark limiter'' purported to be an improvement on previous bark detectors (such as covered by United States Patent 4,947,795). An electronic circuit applies electrical stimuli to the dog in order to discourage the animal from barking further. The object of the detector is to discriminate between various sounds and vibrations that a dog might make and avoid applying the deterring electrical stimuli to any sound other than a valid bark sound.—DRR



6,928,769

43.80.Nd DISPOSABLE INSECT-CONTROL MEMBER

Lantz S. Crawley and J. Roy Nelson, assignors to Bugjammer, Incorporated
16 August 2005 (Class 43/107); filed 7 August 2001

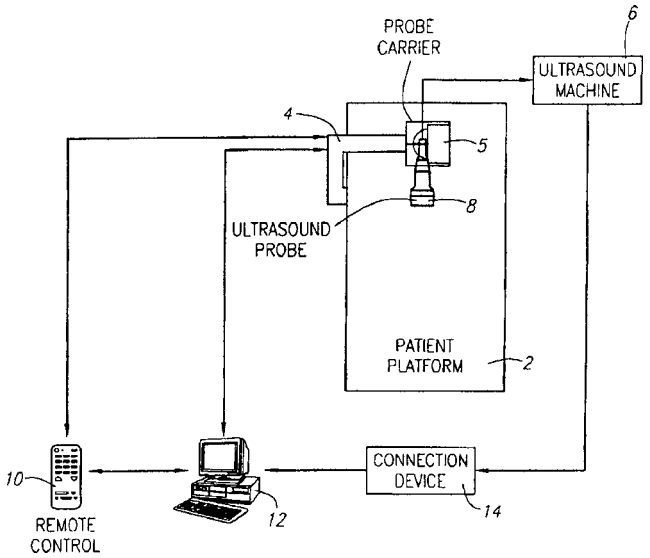
In a preferred embodiment, this device consists of a flexible or semi-rigid vibration-coupling surface facing an insect-engagement surface. The vibration-coupling surface attaches to a mechanical driver so that it vibrates in a prescribed frequency pattern, one that mimics the heartbeat of an animal, in order to lure flying insects to the vicinity of the insect-engagement surface. This surface serves as a backing for an insect-trapping material, such as an adhesive, pesticide, or mineral oil.—DRR

6,932,768

43.80.Qf ULTRASONIC CELLULAR TISSUE SCREENING SYSTEM

Kevin M. Kelly *et al.*, assignors to Sonocine, Incorporated
23 August 2005 (Class 600/437); filed 21 March 2003

The goal of this ultrasound system is to screen cellular tissue, in particular, breast tissue. The system employs an ultrasound probe for generating image data representing the cellular tissue and one or more sensors to determine the probe's location. An ultrasonically conductive pad is placed over the patient's nipple and a fabric covering is placed over the breast tissue to



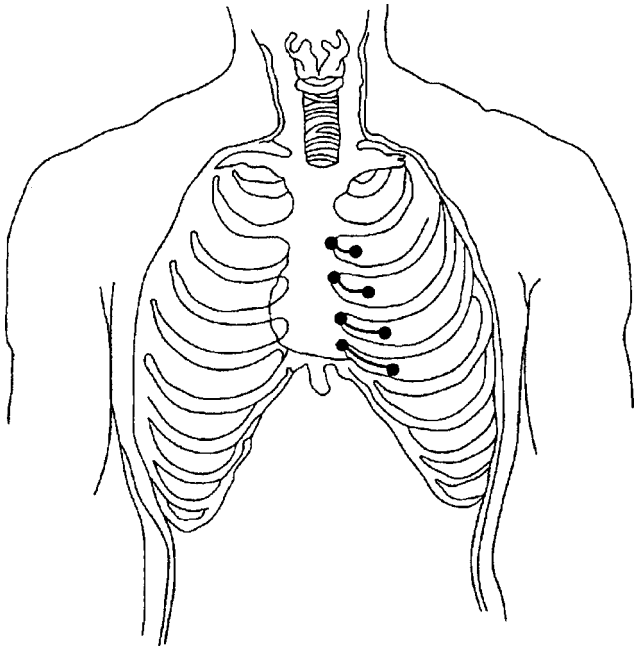
be scanned. The pad possesses ultrasonic characteristics different from those of the cellular tissue to be scanned. The fabric covering contains an ultrasonic coupling agent and is used to hold the breast tissue in place. Thus, ultrasound energy is transmitted with minimal interference. The images representing the scanned tissue are displayed on a viewer capable of providing a rapid sequence of images.—DRR

6,939,308

43.80.Qf ACOUSTIC SENSOR ARRAY FOR NON-INVASIVE DETECTION OF CORONARY ARTERY DISEASE

Charles E. Chassaing and Hung Nguyen, assignors to MedAcoustics, Incorporated
6 September 2005 (Class 600/504); filed 8 November 2002

This device takes advantage of the fact that an ''acoustic window'' can be identified as that area above the notch in the human left lung which allows the heart to be in contact with the chest wall. A sensor array positioned on a patient's chest substantially within the perimeter of this acoustic window can generate well-correlated acoustic blood flow signals. The



device consists of sensor arrays located within this acoustic window and incorporates means of identifying the acoustic window from the merged window subareas corresponding to two or more intercostal spaces. In this manner, the most revealing data regarding cardiac performance can be obtained.—DRR

6,934,579

43.80.Qf ANAESTHESIA CONTROL SYSTEM

Haralambos Mantzaridis and Gavin N. C. Kenny, assignors to The University Court of the University of Glasgow
23 August 2005 (Class 600/544); filed in the United Kingdom 11 September 1996

This system subjects the patient to repetitive audio stimuli, monitors auditory evoked potentials (AEPs) produced by the patient, and then records these AEPs through the use of EEG recording. A signal corresponding to the coarseness of the monitored AEP signal is used as an indicator of anaesthetic depth. The raw AEP signal is divided into a series of sweeps, each synchronized with the repetitive audio stimulus in order to produce a time-averaged sweep from which the anaesthetic index is calculated. This system and index signal can be used as a part of the procedure to regulate the anaesthetic supply to the patient, maintaining the anaesthetic index at a predetermined level.—DRR

6,929,632

43.80.Sh ULTRASONIC DEVICES AND METHODS FOR ABLATING AND REMOVING OBSTRUCTIVE MATTER FROM ANATOMICAL PASSAGEWAYS AND BLOOD VESSELS

Henry Nita and Timothy C. Mills, assignors to Advanced Cardiovascular Systems, Incorporated
16 August 2005 (Class 604/508); filed 27 June 2002

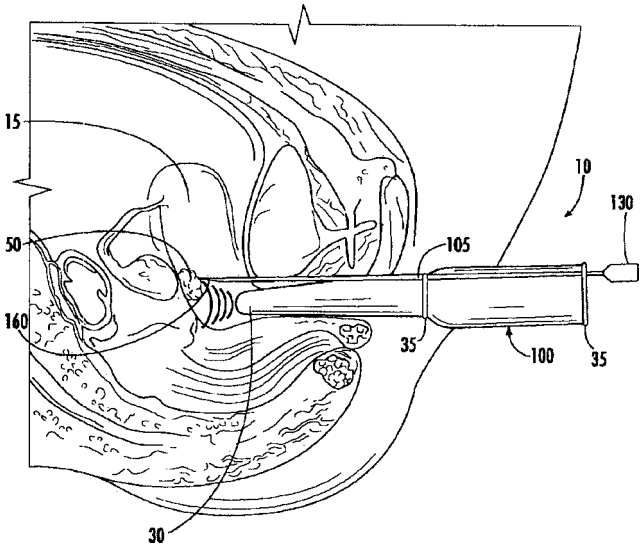
An ultrasound delivery catheter incorporates one or more aspiration lumens extending longitudinally through the catheter to aspirate and remove particles or other debris from an anatomical structure in which an obstruction is located. Potential applications include ultrasonic treatment of obstructions within intracranial and extracranial cerebral blood vessels and removal of obstructions in the male or female urogenital tracts, such as those in fallopian tubes and prostatic areas.—DRR

6,936,048

43.80.Sh ECHOGENIC NEEDLE FOR TRANSVAGINAL ULTRASOUND DIRECTED REDUCTION OF UTERINE FIBROIDS AND AN ASSOCIATED METHOD

Bradley Shawn Hurst, assignor to Charlotte-Mecklenburg Hospital Authority
30 August 2005 (Class 606/41); filed 16 January 2003

This device is essentially a medical needle for transvaginal ultrasound-directed reduction of fibroids. The echogenic needle, configured for use in conjunction with a transvaginal ultrasound probe, has an echogenic surface near its tip that allows the physician to visualize its location in the ultrasound image. In one embodiment, the needle has an active electrode at its



distal end, which supplies rf energy to a fibroid, thereby causing necrosis of the targeted fibroid or destroying the fibroid's vascular supply. The rf needle may also have a safety device that shuts off the energy if the needle penetrates the uterine wall. Another embodiment employs a cryogen supply to destroy a fibroid or its vascular supply by freezing one or the other with the needle's frozen distal end.—DRR

6,936,008

43.80.Vj ULTRASOUND SYSTEM WITH CABLELESS COUPLING ASSEMBLY

Umit Tarakci *et al.*, assignors to Zonare Medical Systems, Incorporated
30 August 2005 (Class 600/437); filed 20 October 2001

A two-dimensional array of ultrasound transducers in this system is wirelessly coupled to the imaging electronics and can also be coupled to a motherboard through a connector with a high density of pins.—RCW

6,936,009

43.80.Vj MATCHING LAYER HAVING GRADIENT IN IMPEDANCE FOR ULTRASOUND TRANSDUCERS

Venkat Subramaniam Venkataramani *et al.*, assignors to General Electric Company
30 August 2005 (Class 600/459); filed 27 February 2001

An impedance gradient is produced by composite-material layers that are bonded together. The layer adjacent to the transducer has an impedance less than or equal to the transducer material. The layer adjacent to the target has an impedance greater than or equal to that of the target. The impedance values decrease monotonically from the transducer to the target.—RCW

6,939,302

43.80.Vj APPARATUS AND METHOD FOR CONTROLLING CONTRAST ENHANCED IMAGING PROCEDURES

David M. Griffiths and Arthur E. Uber III, assignors to MEDRAD, Incorporated
6 September 2005 (Class 600/458); filed 19 October 2001

The concentration of an ultrasound contrast enhancement agent is measured by a concentration sensor during injection that is performed by a device that applies pressure to move the agent through a fluid path to the site of interest.—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.

A numerically accurate and robust expression for bistatic scattering from a plane triangular facet (L)

Gorm Wendelboe[a)]
Danish Defense Research Establishment (DDRE), Ryvangs Allé 1, DK-2100 Copenhagen and Ørsted-DTU, Technical University of Denmark, Ørsteds Plads, Building 348, DK-2800 Kgs. Lyngby, Denmark

Finn Jacobsen[b)]
Ørsted-DTU, Technical University of Denmark, Ørsteds Plads, Building 348, DK-2800 Kgs. Lyngby, Denmark

Judith M. Bell[c)]
School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, United Kingdom

(Received 18 March 2005; revised 22 July 2005; accepted 16 November 2005)

This work is related to modeling of synthetic sonar images of naval mines or other objects. Considered here is the computation of high frequency scattering from the surface of a rigid 3D-object numerically represented by plane triangular facets. The far field scattered pressure from each facet is found by application of the Kirchhoff approximation. Fawcett [J. Acoust. Soc. Am. **109**, 1319–1320 (2001)] derived a time domain expression for the backscattered pressure from a triangular facet, but the expression encountered numerical problems at certain angles, and therefore, the effective ensonified area was applied instead. The effective ensonified area solution is exact at normal incidence, but at other angles, where singularities also exist, the scattered pressure will be incorrect. This paper presents a frequency domain expression generalized to bistatic scattering written in terms of sinc functions; it is shown that the expression improves the computational accuracy without loss of robustness. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2149842]

PACS number(s): 43.20.Fn, 43.20.Px [MO] Pages: 701–704

I. INTRODUCTION

In the past decade low price hardware has made high frequency sonar imagery systems widely available for naval-mine identification, harbor surveillance, and offshore industry. This work is related to the generation synthetic sonar images of naval mines and other objects.

An object will be numerically represented by elementary facets. The far field scattered pressure from each facet is found by application of the physical optics solution or Kirchhoff approximation.[1,2] The total scattered field is given as the coherent sum of pressure contributions from all nonshadowed facets.

The plane rectangular facet has been used by Sammelmann[3] and George.[4] However, for arbitrarily shaped objects the plane rectangular facet can lead to problems concerning correct surface representation. The plane triangular facet, on the other hand, is suited for all types of surfaces because of its co-planar property. Fawcett[5] derived the time domain impulse response for backscattering from a plane triangular facet. An alternative to the flat facet approach is the application of nonuniform rational B-spline surfaces (NURBS).[6] The field integral is evaluated over a parametric space of Bezier surfaces using the method of stationary phase.

The plane triangular facet is considered in this note. The expression for the scattered pressure from the plane triangular facet, first presented by Fawcett,[5] consists of three contributions corresponding to a response from each of the three vertices. However, this vertex response, which is applicable in the time domain, becomes numerically unstable at certain angles as the angle dependent coefficients get very large due to division by very small numbers. One way to deal with this is to replace the integral solution with the effective ensonified area, when the coefficients exceed, say, 1000.[5] This

[a)]Electronic mail: gw@oersted.dtu.dk
[b)]Electronic mail: fja@oersted.dtu.dk
[c)]Electronic mail: j.bell@hw.ac.uk

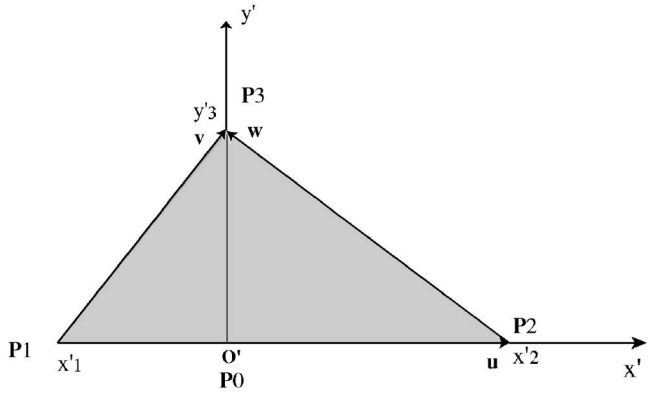


FIG. 1. Local coordinate system ($'$) of a plane triangular facet.

choice might be appropriate for near normal incidence, but for angles away from near normal incidence, where singularities also exist, the scattered pressure will be incorrect. In this note the time domain opportunity is abandoned and the frequency domain expression is rewritten in a numerically robust frame formulated additionally for bistatic scattering.

II. FACET-GEOMETRY AND FIELD APPROXIMATIONS

In this section a scattering integral for a rigid, plane and triangular facet is derived. The Kirchhoff approximation as well as the far field approximation are applied.

Suppose an arbitrarily shaped body is numerically represented by plane triangular facets. Each facet in the three-dimensional space is represented by its vertex points $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ and the unit surface normal vector, $\hat{\mathbf{n}}_s$, pointing out of the body. The vectors connecting the vertex points are defined by $\mathbf{u}=\overrightarrow{\mathbf{P}_1\mathbf{P}_2}$, $\mathbf{v}=\overrightarrow{\mathbf{P}_1\mathbf{P}_3}$, and $\mathbf{w}=\overrightarrow{\mathbf{P}_2\mathbf{P}_3}$, and they are arranged such that $\mathbf{u}$ represents the longest side of the triangle and $(\mathbf{u}\times\mathbf{v})\cdot\hat{\mathbf{n}}_s>0$; see Fig. 1.

A local coordinate system ($'$) is introduced. The origin $\mathbf{O}'$ has the global coordinates $\mathbf{P}_0=\mathbf{P}_1+\mathbf{v}_\mathbf{u}$, where $\mathbf{v}_\mathbf{u}$ is the projection of $\mathbf{v}$ on $\mathbf{u}$. In the local coordinate system the triangle is described by the axis-points x_1', x_2', and y_3'. The base of ($'$) is

$$\mathbf{e}_x' = \mathbf{u}/|\mathbf{u}|, \tag{1a}$$

$$\mathbf{e}_y' = (\mathbf{v}-\mathbf{v}_\mathbf{u})/|\mathbf{v}-\mathbf{v}_\mathbf{u}|, \tag{1b}$$

$$\mathbf{e}_z' = \mathbf{e}_x' \times \mathbf{e}_y'. \tag{1c}$$

The base given by Eqs. (1a)–(1c) establishes a coordinate transformation matrix applied on $\mathbf{r}_0$, the vector from $\mathbf{P}_0$ to the source, and on $\mathbf{r}_1$, the vector from $\mathbf{P}_0$ to the observation point. In the ($'$)-coordinate system $\mathbf{r}_0$ and $\mathbf{r}_1$ have been transformed into $\mathbf{r}_0'$ and $\mathbf{r}_1'$, respectively, but their (Euclidian) lengths are unchanged, i.e., $|\mathbf{r}_0'|=|\mathbf{r}_0|=r_0$ and $|\mathbf{r}_1'|=|\mathbf{r}_1|=r_1$. The unit surface normal vector has been transformed into $\hat{\mathbf{n}}_s' \equiv [001]^T$, and each point on the plane facet, S, is given by $\mathbf{x}_s' \equiv [x'y'0]^T$; see Fig. 2.

The Kirchhoff approximation is applied on the rigid surface and the total pressure field, p, is related to the incoming field, p_{inc}, through

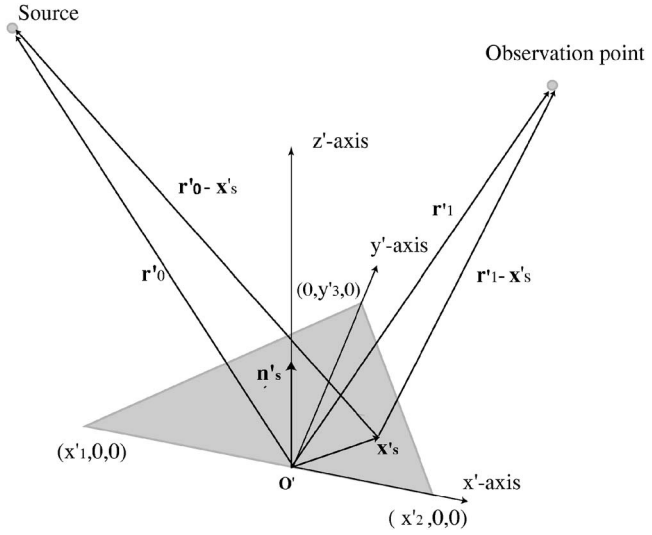


FIG. 2. Geometry applied for field integration over the surface of a triangular facet.

$$p(\mathbf{x}_s') = 2p_{\text{inc}}(\mathbf{x}_s') \tag{2}$$

(see, e.g., Fawcett[5]). The Kirchhoff Helmholtz integral equation [see, e.g., Pierce,[7] Eq. (4-6.4)] consequently reduces to the integral

$$p_{\text{sc}}(\mathbf{r}_1') = \frac{1}{2\pi}\int_S p_{\text{inc}}(\mathbf{x}_s')\,\nabla\, G(\mathbf{x}_s'|\mathbf{r}_1')\cdot\hat{\mathbf{n}}_s'\,dS, \tag{3}$$

where $p_{\text{sc}}(\mathbf{r}_1')$ is the facet-scattered pressure measured at the observation point, $\mathbf{r}_1'$, and $\nabla G(\mathbf{r}_1'|\mathbf{x}_s')$ is the gradient of the free space Green's function directed toward the observation point. The incoming field originates from a monopole source,

$$p_{\text{inc}}(\mathbf{x}_s') = p_0\frac{e^{ik|\mathbf{x}_s'-\mathbf{r}_0'|}}{|\mathbf{x}_s'-\mathbf{r}_0'|}, \tag{4}$$

where p_0 is the pressure amplitude 1 meter from the source. In the far field, the range from the facet to the source by far exceeds the dimensions of the facet, and hence, a first order Taylor series expansion of $|\mathbf{x}_s'-\mathbf{r}_0'|$ with respect to $\mathbf{x}_s'/r_0$ is possible,

$$|\mathbf{x}_s'-\mathbf{r}_0'| \approx r_0 - \hat{\mathbf{r}}_0'\cdot\mathbf{x}_s', \tag{5}$$

where $\hat{\mathbf{r}}_0'=\mathbf{r}_0'/r_0$ [see, e.g., Ogilvy,[1] Eq. (4.10)]. The right-hand side of Eq. (5) will be applied for the phase term of Eq. (4) whereas it is sufficient to approximate the slowly varying denominator in Eq. (4), i.e., the term that represents geometrical spreading, to zero order, that is, $|\mathbf{x}_s'-\mathbf{r}_0'|\approx r_0'$. The incoming field is

$$p_{\text{inc}}(\mathbf{x}_s') = p_0\frac{e^{ikr_0}}{r_0}e^{-ik\hat{\mathbf{r}}_0'\cdot\mathbf{x}_s'}, \tag{6}$$

and hence, the part of the spherical wave that sweeps over the facet is considered locally plane. The gradient of the Green's function,

$$\nabla G(\mathbf{r}_1'|\mathbf{x}_s') = \frac{\mathbf{r}_1' - \mathbf{x}_s'}{|\mathbf{r}_1' - \mathbf{x}_s|^3}(ik|\mathbf{r}_1' - \mathbf{x}_s| - 1)e^{ik|\mathbf{r}_1' - \mathbf{x}_s'|}, \tag{7}$$

[see, e.g., Pierce,[7] Eq. (4-6.5)] is also approximated to the far field. In the phase term, $|\mathbf{r}_1' - \mathbf{x}_s'| \approx r_1 - \hat{\mathbf{r}}_1' \cdot \mathbf{x}_s'$, is applied, and in the amplitude terms, $\mathbf{r}_1' - \mathbf{x}_s' \approx \mathbf{r}_1'$ is used. The expression within the brackets of Eq. (7) is approximated to $(ik|\mathbf{r}_1' - \mathbf{x}_s'| - 1) \approx (ikr_1 - 1) \approx ikr_1$, where the second approximation is valid because $ikr_1 \gg 1$. Hence,

$$\nabla G(\mathbf{r}_1'|\mathbf{x}_s') = \hat{\mathbf{r}}_1' ik \frac{e^{ikr_1}}{r_1} e^{-ik\hat{\mathbf{r}}_1' \cdot \mathbf{x}_s'}, \tag{8}$$

where $\hat{\mathbf{r}}_1' = \mathbf{r}_1'/r_1$. Insertion of Eq. (6) and Eq. (7) into Eq. (3) leads to a scattering integral with a linear phase term

$$p_{\mathrm{sc}}(\mathbf{x}) = \frac{p_0 ike^{ik(r_0+r_1)}\hat{\mathbf{r}}_1' \cdot \hat{\mathbf{n}}_s'}{2\pi r_0 r_1} \int_S e^{-ik(\hat{\mathbf{r}}_0' + \hat{\mathbf{r}}_1') \cdot \mathbf{x}_s'} \, dS. \tag{9}$$

III. A ROBUST EXPRESSION FOR TRIANGULAR FACET SCATTERING

If the phase variation over the facet is neglected, the integral in Eq. (9) can be replaced by the facet area, S, and the approximate effective ensonified area response is

$$p_{\mathrm{sc}}(\mathbf{x}) = \frac{p_0 ike^{ik(r_0+r_1)} \cos\theta_1}{2\pi r_0 r_1} S, \tag{10}$$

where $\cos\theta_1 = \hat{\mathbf{r}}_1' \cdot \hat{\mathbf{n}}_s'$. However, this expression is only exact for normal incidence. In the general case the dot-product in the exponential term of the surface integral Eq. (9) must be considered,

$$(\hat{\mathbf{r}}_0' + \hat{\mathbf{r}}_1') \cdot \mathbf{x}_s' = ax' + by', \tag{11}$$

where the angle dependent constants are given by

$$a = \sin\theta_0 \cos\varphi_0 + \sin\theta_1 \cos\varphi_1, \tag{12a}$$

$$b = \sin\theta_0 \sin\varphi_0 + \sin\theta_1 \sin\varphi_1, \tag{12b}$$

and where the angle between $\mathbf{r}_j'$ and $\mathbf{n}_s'$ is $0 \leqslant \theta_j \leqslant \pi/2$, and the angle in the $x'y'$ plane is $0 \leqslant \varphi_j \leqslant 2\pi$, $j = 0, 1$. The solution to Eq. (9) is brought on the vertex response form

$$p_{\mathrm{sc}}(\mathbf{x}) = \frac{p_0 e^{ik(r_0+r_1)} \cos\theta_1}{ik2\pi r_0 r_1}[\kappa_1 e^{-ikax_1'} + \kappa_2 e^{-ikax_2'} + \kappa_3 e^{-ikby_3'}], \tag{13a}$$

where κ_1, κ_1, and κ_3 are the vertex coefficients given by

$$\kappa_1 = \frac{-y_3'}{a(ax_1' - by_3')}, \tag{13b}$$

$$\kappa_2 = \frac{y_3'}{a(ax_2' - by_3')}, \tag{13c}$$

$$\kappa_3 = \frac{y_3'(x_2' - x_1')}{(ax_2' - by_3')(ax_1' - by_3')}. \tag{13d}$$

The vertex response can be transformed into the time domain to obtain the impulse response (see, e.g., Fawcett[5]). As can be observed from Eqs. (13b)–(13d) three singularities are present, $a = 0$, $ax_1' - by_3' = 0$, and $ax_2' - by_3' = 0$. For backscattering these singularities correspond to incident directions normal to the three sides of the triangle, i.e., normal to **u**, **w** or **v**. From an analytical point of view, large values of κ_j will cancel each other in Eq. (13a). However, in a numerical implementation the cancellation tends to fail because of truncation errors obtained near the working precision of the computer. In what follows the time domain approach is abandoned and a numerically robust expression is derived. The angle dependent terms a, $(ax_1' - by_3')$, and $(ax_2' - by_3')$ are separated and expressed in terms of the well-behaved sinc function. Hence, the solution to Eq. (9) is written

$$\begin{aligned} p_{\mathrm{sc}}(\mathbf{x}) = {} & \frac{p_0 e^{ik(r_0+r_1)} \cos\theta_1}{2\pi r_0 r_1}\left(\frac{1}{ikb}\right) \\ & \times \left[x_1' e^{-ik(ax_1' + by_3')/2} \frac{\sin(k[ax_1' - by_3']/2)}{k[ax_1' - by_3']/2} \right. \\ & - x_2' e^{-ik(ax_2' + by_3')/2} \frac{\sin(k[ax_2' - by_3']/2)}{k[ax_2' - by_3']/2} \\ & \left. + (x_2' - x_1') e^{-ika(x_1' + x_2')/2} \frac{\sin(ka[x_2' - x_1']/2)}{ka[x_2' - x_1']/2} \right]. \end{aligned} \tag{14a}$$

When $b \to 0$, Eq. (14a) becomes numerically unstable, and is replaced by the limit value, $p_{\mathrm{sc}}(\mathbf{x})$ for $b \to 0$, found by using the rule of L'Hospital,

$$\begin{aligned} p_{\mathrm{sc}}(\mathbf{x}) = {} & \frac{p_0 e^{ik(r_0+r_1)} \cos\theta_1 y_3'}{4\pi i r_0 r_1} \times [g(kax_1'/2)x_1' e^{-ikax_1'/2} \\ & - g(kax_2'/2)x_2' e^{-ikax_2'/2}], \end{aligned} \tag{14b}$$

where

$$g(x) = \frac{-\cos(x) + \sin(x)/x}{x} - i\frac{\sin(x)}{x}. \tag{14c}$$

IV. RESULTS

Validation is carried out by considering the canonical problem of backscattering of a plane wave of unit amplitude from a rigid sphere. The infinite harmonic series solution[8] is the Benchmark solution. The sphere is a good test case because sharp edges, where the Kirchhoff approximation fails, are absent. The time domain response of a 445 kHz Ricker pulse incident on a sphere with a radius equal to 10 centimeters measured at a range of 10 meters is determined. In the numerical model the surface of the sphere is represented by iso-sized facets with areas of 11 mm^2. Results from three different numerical Kirchhoff approaches are considered. The first approach is the vertex response given by Eqs. (13) where no numerical precautions are taken. In the second approach the vertex response is combined with the effective area response given by Eq. (10) according to Fawcett.[5] The vertex response is switched "off" and the effective area response "on" when the magnitude of one of the coefficients in Eqs. (13) exceed 1000. Finally, the third ap-

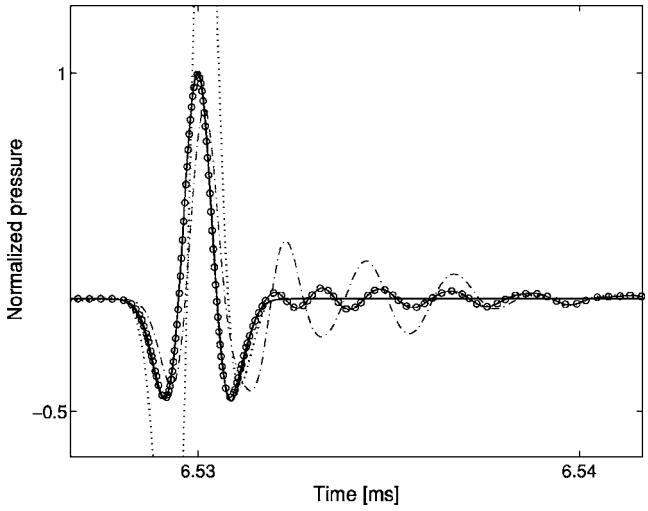


FIG. 3. Backscattering from a rigid sphere. The thick solid line is the benchmark solution; the dotted line is the vertex response [Eq. (13)] the dashed-dotted line is the vertex response, [Eq. (13)], combined with the effective area response [Eq. (10)]; and the solid line with circles is the robust response [Eq. (14)].

proach is the robust response given by Eq. (14). Fourier synthesis has been carried out on the benchmark solution as well as on the three different numerical Kirchhoff approaches.

The results are given in Fig. 3. The robust response given by Eq. (14) matches the analytic solution very well except after the specular reflection where oscillations occur. These oscillations are related to the numerical surface discretization of the sphere and not the scattering formulas. The result from the pure vertex response, Eqs. (13), clearly diverges from the analytical solution at the specular reflection, but after that, it quickly approaches the robust response. Hence, the pure vertex response is accurate as long as the critical angles are not encountered; if this happens, erroneous results orders of magnitudes from the actual response will occur. The combined solution is significantly closer to the analytical result at the specular reflection, but exhibits stronger oscillations after the specular reflection, indicating that the combined approach is a robust, but not very accurate, numerical approach when singularities are encountered. The robust response expressed in terms of sinc functions has proven to be the most numerically reliable expression.

V. CONCLUSION

A numerically robust expression for the far field bistatic scattered pressure from a plane triangular facet written in terms of sinc functions, Eq. (14), has been presented. The expression is applied for the computation of high frequency scattering from arbitrarily shaped objects. Equation (14) has been compared with two solutions based on the vertex response expression, Eq. (13). The first solution is the pure vertex response which is numerically unstable at certain angles. The second solution, presented by Fawcett,[5] combines Eq. (13) with the effective ensonified area, Eq. (10), in the case where one of the coefficients of the vertex response, Eqs. (13b)–(13d), exceeds a threshold value equal to 1000. Comparisons have been carried out by considering the canonical problem of backscattering of a plane wave from a rigid sphere, and the expression written in terms of sinc functions has been shown to be the most reliable solution in terms of accuracy.

ACKNOWLEDGMENT

The authors gratefully acknowledge communication with John Fawcett at DRDC, Atlantic.

[1] J. A. Ogilvy, *Theory of Wave Scattering from Random Rough Surfaces* (IOP, London, 1991), Chap. 4.1.
[2] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic Press, London, 1998), Chap. 7.2.3.
[3] G. S. Sammelman, "Propagation and scattering in very shallow water," Proc. IEEE Oceans **1**, 337–344 (2001).
[4] O. George and R. Bahl, "Simulation of backscattering of high frequency sound from complex objects and sand sea-bottom," IEEE J. Ocean. Eng. **20**, 119–130 (1995).
[5] J. A. Fawcett, "Modeling of high-frequency scattering from objects using a hybrid Kirchhoff/diffraction approach," J. Acoust. Soc. Am. **109**, 1312–1319 (2001).
[6] J. Perez and M. F. Catedra, "Application of physical optics to the RCS computation of bodies modeled with NURBS surfaces," IEEE Trans. Antennas Propag. **42**, 1404–1411 (1994).
[7] A. D. Pierce, *Acoustics. An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1991), second printing.
[8] E. G. Williams, *Fourier Acoustics. Sound Radiation and Nearfield Acoustical Holography* (Academic Press, London, 1999), Eq. (6.185).

Geometric sound propagation through an inhomogeneous and moving ocean: Scattering by small scale internal wave currents (L)

John A. Colosi[a)]
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

(Received 20 December 2004; revised 22 November 2005; accepted 29 November 2005)

Ray equations appropriate for ocean acoustic propagation through an inhomogeneous and moving ocean are put forth with applications to sound scattering by internal waves. The result reveals the important role played by range dependent horizontal current shear. The Garrett–Munk internal wave model and observations of upper ocean shear and sound speed fluctuations suggest that inertial frequency upper ocean shear may play a comparable role to internal wave induced sound speed fluctuations as a source of upper ocean acoustic scattering. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159587]

PACS number(s): 43.30.Cq, 43.30.Ft, 43.30.Es [DRD] Pages: 705–708

I. INTRODUCTION

The physics of infragravity and acoustic wave propagation in the atmosphere is strongly influenced by the significant shear that exists in the air (Lighthill 1978; Pierce 1989). In ocean acoustic propagation however, the effects of the relatively gentle ocean currents have been mostly of interest in acoustic travel time tomography using reciprocal transmissions (Munk, Worcester, and Wunsch 1995 and references therein). There have been a few examples of theoretical and observational efforts to measure current effects from small scale ocean processes like internal waves (Worcester 1977; Munk *et al.* 1981; Stoughton *et al.* 1986), but here weak scattering and travel time fluctuations were the focus.

While some general treatments exist (Uginčius 1965, 1972;Thompson 1972; Godin and Voronovich, 2004), the ocean acoustic literature on the theory of the effects of currents on sound propagation has somewhat overemphasized cases in which the current is only a function of depth (Blokhintzev 1946; Keller 1954; Stallworth and Jacobson 1972a, b; Franchi and Jacobson 1972; Robertson *et al.* 1985) or only the observable of travel time is treated in detail (Munk, Worcester, and Wunsch 1995). In the theory of ocean acoustic scattering, established work in this area (Flatte *et al.* 1979; Colosi *et al.* 1999 and references therein) has not rigorously evaluated the effects of ocean currents on acoustic propagation, and the focus has been almost entirely on sound speed fluctuations from the Garrett-Munk internal wave model (Munk 1981). With the recent work by Dzieciuch, Munk, and Rudnick (2004), who point out acoustic scattering mechanisms by upper ocean density compensated intrusive features, it is reasonable to reevaluate the scattering mechanism from upper ocean currents.

The purpose of this paper is to examine general ray-equations for two-dimensional (2-D) ocean acoustic propagation through range dependent sound speed and current fields, and to provide a discussion of the relative effects of sound speed and current fluctuations with regard to acoustic scattering. The ray equations are presented in such a way as to reveal the important effects of current shear which are not apparent in other approaches (Uginčius 1965, 1972; Thompson 1972; Godin and Voronovich, 2004). It is shown using the Garrett-Munk (GM) internal wave model (Munk 1981) and observed fields of upper ocean current and sound speed fluctuation that the current shear effects may in fact be significant in distorting the acoustic ray path. Another recent paper with a slightly different approach comes to the same conclusion (Duda 2005). Further analysis, which may include direct numerical evaluations of the equations of motion or more theoretical work, both which are beyond the scope of this paper, will be needed to resolve this issue.

II. 2-D GEOMETRICAL ACOUSTICS

To address a rather simple case, spatially varying but time frozen fields of sound speed and current of the form $c = c(x,z)$ and $\vec{v} = [u(x,z),0,0]$ are assumed where x is the horizontal coordinate and z is the vertical (depth) coordinate. The generalization to three-dimensional cases is algebraically cumbersome but straight forward. The vertical component of the current is neglected since in a vertically stratified ocean the horizontal component is much larger than the vertical component (Pedlosky 1986). Again, incorporation of all current components is straightforward, but algebraically tedious without providing much more insight. In this paper we only consider the effects of advection on the wave propagation (See Pierce 1989, page 402 for other mechanisms) and thus we write the wave equation for pressure p using the advective time derivative so that

$$c^2(\partial_{xx} + \partial_{zz})p = (\partial_t + u\partial_x)^2 p. \tag{1}$$

The acoustic dispersion relation for Eq. (1) is then given by

$$\omega(k_x,k_z,x,z) = c(x,z)(k_x^2 + k_z^2)^{1/2} + u(x,z)k_x \tag{2}$$

and thus the rays are easily obtained by well-known methods (See Lighthill 1978). The ray equations are:

[a)]Present address: Naval Postgraduate School, Department of Oceanography, Monterey CA

$$\frac{dk_x}{\partial t} = -\frac{\partial \omega}{\partial x} = -k_x \frac{\partial u}{\partial x} - (k_x^2 + k_z^2)^{1/2} \frac{\partial c}{\partial x} \tag{3}$$

$$\frac{dk_z}{\partial t} = -\frac{\partial \omega}{\partial z} = -k_x \frac{\partial u}{\partial z} - (k_x^2 + k_z^2)^{1/2} \frac{\partial c}{\partial z} \tag{4}$$

$$\frac{dx}{dt} = \frac{\partial \omega}{\partial k_x} = u + c \frac{k_x}{(k_x^2 + k_z^2)^{1/2}} \tag{5}$$

$$\frac{dz}{dt} = \frac{\partial \omega}{\partial k_z} = c \frac{k_z}{(k_x^2 + k_z^2)^{1/2}} \tag{6}$$

These equations can be manipulated to give the ray slope

$$\frac{dz}{dx} = \frac{dz/dt}{dx/dt} = \frac{ck_z}{u(k_x^2 + k_z^2)^{1/2} + ck_x} \equiv \tan\,\theta. \tag{7}$$

In a stationary medium where c and u are not explicit functions of time the frequency ω is conserved along the ray path (Lighthill 1978) and thus the variable k_x can be solved for in terms of the frequency yielding

$$k_x = \frac{-\omega u + c\sqrt{\omega^2 - (c^2 - u^2)k_z^2}}{(c^2 - u^2)}. \tag{8}$$

Here the positive root is taken for propagation in the positive x direction. Substituting this result into our equation for ray slope gives

$$\frac{dz}{dx} = \frac{k_z c}{\sqrt{\omega^2 - k_z^2(c^2 - u^2)}}. \tag{9}$$

Next an equation for k_z is needed, which comes readily from the ray equations

$$\frac{dk_z}{dx} = -\frac{\partial \omega}{\partial t}\frac{dt}{dx} = \frac{k_x u - \omega}{\sqrt{\omega^2 - k_z^2(c^2 - u^2)}} \times \left(\frac{k_x}{c}\frac{\partial u}{\partial z} + (\omega - k_x u)\frac{1}{c^2}\frac{\partial c}{\partial z}\right) \tag{10}$$

and here the dependence on k_x has been retained since further simplification is not possible. Finally a result for ray travel time is needed. The result comes, again, directly from the ray equations and with a little manipulation we obtain,

$$\frac{dT}{dx} = \frac{1}{dx/dt} = \frac{1}{(c^2 - u^2)}\left(-u + \frac{c}{\sqrt{\omega^2 - k_z^2(c^2 - u^2)}}\right). \tag{11}$$

Since ω is conserved along a ray, and since k_x and k_z are both proportional to ω we can scale out this factor by introducing the "slowness" coordinates $p_x = k_x/\omega$ and $p_z = k_z/\omega$. The ray equations then become

$$\frac{dz}{dx} = \frac{p_z c}{\sqrt{1 - p_z^2(c^2 - u^2)}}, \tag{12}$$

$$\frac{dp_z}{dx} = \frac{p_x u - 1}{\sqrt{1 - p_z^2(c^2 - u^2)}}\left(\frac{p_x}{c}\frac{\partial u}{\partial z} + (1 - p_x u)\frac{1}{c^2}\frac{\partial c}{\partial z}\right), \tag{13}$$

$$\frac{dT}{dx} = \frac{1}{(c^2 - u^2)}\left(-u + \frac{c}{\sqrt{\omega^2 - p_z^2(c^2 - u^2)}}\right). \tag{14}$$

Equations (12)–(14) are the basic equations describing the geometry and phase (travel time) of the acoustic ray paths, and in this form can be easily solved on the computer. Note the similar role that current shear (du/dz) and sound speed gradient (dc/dz) play in Eq. (13), a result that is not as easily apparent on other general treatments (Uginčius 1965, 1972; Thompson 1972; Godin and Voronovich, 2004). Equations (12) and (13) of course reduce to the standard results when $u=0$ (Brown *et al.* 2003).

An understanding of the travel time Eq. (14) can be obtained by writing it in terms of the ray angle θ. Equation (12) is equal to the tangent of the ray angle ($\tan\,\theta$) therefore we can solve this equation giving $p_z = \sin\,\theta/\sqrt{c^2 - u^2\sin^2\,\theta}$ (note for $u=0$ the standard result $p_z = \sin\,\theta/c$ is obtained). Substituting this expression for p_z into Eq. (14) gives

$$\frac{dT}{dx} = \frac{\sec\,\theta}{c^2 - u^2}(-u\cos\,\theta + \sqrt{c^2 - u^2\sin^2\,\theta}) = \frac{\sec\,\theta}{u\cos\,\theta + \sqrt{c^2 - u^2\sin^2\,\theta}}. \tag{15}$$

Equation (15) is consistent with the results of Thompson (1972), Uginčius (1972), and Munk *et al.* (1995), and of course reduces to the standard result for $u=0$.

III. FIRST-ORDER EFFECTS AND CURRENT SHEAR

Equations (12)–(14) are the basic equations describing the geometry and travel time of the acoustic ray paths, but for ocean acoustic applications it is helpful to obtain some first-order results in Mach number $u/c \ll 1$. With some simple manipulation it is found that,

$$\frac{dT}{dx} = \frac{1}{c^2}\left(\frac{c}{\sqrt{1 - p_z^2 c^2}} - u\right), \tag{16}$$

$$\frac{dz}{dx} = \frac{p_z c}{\sqrt{1 - p_z^2 c^2}}, \tag{17}$$

$$\frac{dp_z}{dx} = -\frac{1}{c^2}\left(\frac{\partial c}{\partial z}\frac{1}{\sqrt{1 - p_z^2 c^2}} + \frac{\partial u}{\partial z}\right), \tag{18}$$

and $p_z \simeq \sin\,\theta/c$. In Eq. (18) we have also neglected a small term of the form $(2u/c^3)\partial c/\partial z$. Equation (16) shows that the travel time correction is simply a small current head/tail wind, an effect that has been known for a long time (Stoughton *et al.* 1986). In Eq. (17) it is seen that the ray slope is unaffected to first order in u/c. Equation (18) displays the most interesting effect that the sound speed gradient and the current shear come in at the same order, since the factor $\sqrt{1 - p_z^2 c^2}$ is of order $\cos\,\theta = O(1)$ for rays with angles $\pm 15°$.

The main issue as to the significance of shear to acoustic propagation are the relative size of shear and sound speed gradients. Let $c(x,z) = c_0(z) + \delta c(x,z)$ where $c_0(z)$ is the mean background sound speed profile and $\delta c(x,z)$ is a perturba-

tion. In typical temperate latitudes mean sound speed gradients in the main thermocline are roughly 0.05 to 0.15 s^{-1}. In polar conditions and in the abyssal ocean the gradient is much smaller and is of order the adiabatic gradient 0.01 s^{-1} (Munk *et al.* 1995). For all intents and purposes we assume the mean current shear to be zero, though this is certainly not true near boundary current regions or near strong fronts.

To estimate the contributions to shear and sound speed gradient from fluctuations we look at internal wave effects. For internal wave induced sound speed fluctuations the well-known result is (Flatté *et al.* 1979)

$$\delta c \simeq GN^2(z)\zeta, \quad \text{and } \delta c' \simeq GN^2(z)\zeta', \tag{19}$$

where $G \sim 1500-2500$ (s), ζ is an internal wave induced vertical displacement, $N(z)$ is the buoyancy frequency profile, and primes denote differentiation with respect to depth. Note that the proportionality constant G that is used is smaller than the canonical value of 3750 (s) (Flatté *et al.* 1979) since recent observations of sound speed variance in the North Pacific are more consistent with the smaller value (Xu *et al.* 2005). From the Garrett-Munk (GM) internal wave spectrum (Munk 1981) the mean square vertical strain is estimated to be

$$\langle \zeta'^2 \rangle = \langle \zeta_0^2 \rangle \frac{N_0}{N} \left(\frac{\pi N}{N_0 B} \right)^2 \frac{2j_*^{j_{\max}}}{\pi} \sum_{j=1} \frac{j^2}{j^2 + j_*^2} \simeq 2\pi \frac{\langle \zeta_0^2 \rangle}{B^2} \frac{N}{N_0} j_* j_{\max}, \tag{20}$$

where j is the internal wave vertical mode number, $j_{\max} \gg j_* = 3$, $\langle \zeta_0^2 \rangle$ is $(7.3\ \text{m})^2$ at a depth where $N = N_0 = 3$ cph, and B is a scale depth for the stratification of order 1000 m. Note here that the vertical mode spectrum has been normalized using $\Sigma_{j=1}^{j_{\max}} (j^2 + j_*^2) \simeq \pi/(2j_*)$. The maximum mode number $j_{\max}$ is a small scale cutoff in the spectrum associated with shear instability; typical values are of order 200–300. A similar calculation for the shear but only considering one component gives (see Munk 1981)

$$\langle u'^2 \rangle = 3\pi \frac{\langle \zeta_0^2 \rangle}{B^2} \frac{N^3(z)}{N_0} j_* j_{\max}, \tag{21}$$

where it is found that the Richardson number $Ri = N^2/\langle u'^2 \rangle$ less than roughly 1/4 is associated with shear instability and thus imposes the cutoff $j_{\max}$. The ratio of shear to sound speed fluctuation however is independent of $j_{\max}$ and so it is found that

$$\frac{u'_{\text{rms}}}{\delta c'_{\text{rms}}} = \left(\frac{\langle u'^2 \rangle}{G^2 N^4 \langle \zeta'^2 \rangle} \right)^{1/2} = \frac{(3/2)^{1/2}}{GN(z)}. \tag{22}$$

Plugging in values of $N = 3$, 1, and 0.5 cph, the rms ratio of shear to sound speed gradient is 0.16, 0.47, and 0.93, respectively: Not an insignificant effect. Both the shear and strain statistics are dominated by internal wave frequencies near the inertial frequency $f = 2\Omega \sin$ (latitude) where Ω is the angular velocity of the earths rotation. Given the crude approximation of the inertial peak by the Garrett-Munk (GM) spectrum the shear effect could be significantly larger (Fu 1981; Garrett 2001; Duda 2005).

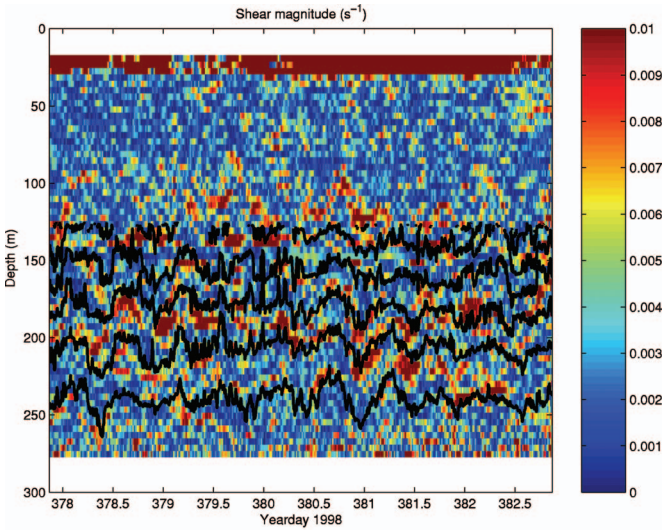


FIG. 1. Shear (color panel) and isotherm depth (black lines) from a mooring with ADCP and temperature, conductivity, and depth (CTD) sensors in the eastern North Pacific Ocean. The mooring was deployed for nearly one year but data shown here were from a five day period January 13 to 18, 1999. Large shears in the shallowest depth regions (20–30 m) are artifacts from ADCP surface interaction, and CTD sensors only extended from 120-m depth to 640-m depth. This image shows that shear in the ocean occurs on relatively thin layers, and that it is advected with the internal wave displacements (isotherms).

The previous discussion is all well and good, but real observations are the best discriminator. Current shears observed in the upper 300 m of the ocean using upward and downward looking 300 kHz acoustic Doppler current profilers (ADCP) are shown in Fig. 1. The data comes from an eastern North Pacific ocean site which was part of the North Pacific acoustic laboratory (NPAL) 1998–1999 field year (See Xu *et al.* 2005 for a description of the data and the experiment). Isotherm depths are also displayed in Fig. 1 to show the advection of the shear along the isotherm surfaces. The shear is concentrated below the mixed layer depth of roughly 100 m, and is associated with the inertial response of the mixed layer to wind forcing (Leaman and Sanford 1975; D'Asaro and Perkins 1984; Pinkel 1984). The current shear has peak magnitudes of roughly 0.005 to 0.01 s^{-1} (Richardson number of order 1; the buoyancy frequency in this depth region is of order 0.0025 to 0.0075 s^{-1} or 1.4 to 4.3 cph (Xu *et al.* 2005)), and a vertical spatial structure which is quite small scale. In addition to ADCP, the NPAL mooring had 16 temperature, conductivity, and pressure sensors spanning a depth region from 150 to 575 m depth which can be used to estimate the sound speed gradients. Figure 2 shows the comparison, and it is evident that both the vertical gradients of sound speed fluctuation ($\delta c'$) and current shear (u') are significantly less than the mean sound speed gradient from the sound channel. Never-the-less the shear effect is comparable to the sound speed fluctuation gradient near 300-m depth, and as expected from the GM theory the sound speed effect becomes more dominant as shallower depths are reached where N is larger. The GM model [Eq. (22)], however, does not get the magnitude of the shear versus sound speed gradient correct as the model predicts a ratio of 1 at roughly 0.5 cph, while the observations in Fig. 2 show a ratio of 1 at about 1.4 cph. This shortcoming

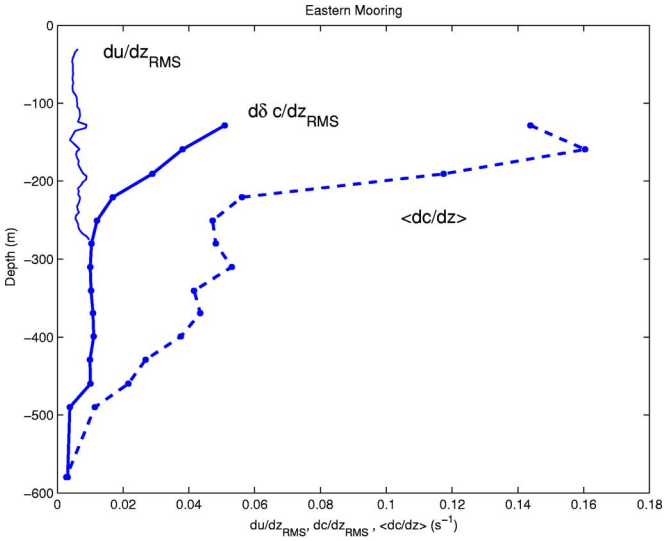


FIG. 2. (Color online) Comparisons of rms shear, rms sound speed gradient, and mean sound speed gradient from a mooring in the eastern North Pacific ocean. Statistical averages were computed between September 1998 and July 1999. In the upper ocean the mean sound speed gradient dominates over the fluctuating sound speed gradients and current shear. The current shear is non-negligible at 250-m depth.

of the GM model is not surprising as upper ocean shear is dominated by inertial waves which are not correctly treated in the GM model. Finally a companion paper to the present work presented by Duda (2005) gives further observational evidence for the comparable effects of shear and strain on ray path propagation.

IV. CONCLUSIONS

The geometrical acoustics equations for propagation through an inhomogeneous and moving medium reveal the important role of current shear in the dynamics. The Garrett-Munk internal wave model and observations of upper ocean shear and sound speed show that shear terms in the ray equations of motion are of comparable magnitude to those terms associated with sound speed gradient. Thus ocean shear may be an important source of scattering in the upper ocean and in the main thermocline, though direct integration of the equations of motion through realistic shear fields will be need to examine the integral effects. This paper suggests that more work needs to be done to examine the effects of currents and shear on sound propagation.

ACKNOWLEDGMENTS

The author would like to acknowledge many useful discussions with Walter Munk and Tim Duda which have provided several improvements to this manuscript. This is Woods Hole Oceanographic Institution contribution number 11276.

Blokhintzev, D. (**1946**), "The propagation of sound in an inhomogeneous and moving medium I," J. Acoust. Soc. Am. **18**(2), 322–328.

Brown, M. G., Colosi, J. A. , Tomsovic, S., Virovlyansky, A. L., Wolfson, M., and Zaslavsky, G. M. (**2003**), "Ray dynamics in long-range deep ocean sound propagation," J. Acoust. Soc. Am. **113**(5), 2533–2547.

Colosi, J. A. (**1999**), "A review of recent results on ocean acoustic wave propagation in random media: Basin scales," IEEE J. Ocean. Eng. **24**(2), 138–155.

D'Asaro, E. A., and Perkins, H. (**1984**), "A near-inertial internal wave spectrum for the Sargasso Sea in late summer," J. Phys. Oceanogr. **25**, 2953–2958.

Duda, T. F. (**2005**), "Ocean sound channel ray path perturbations from internal wave shear and strain," J. Acoust. Soc. Am. **118**(5), 2899–2903.

Dzieciuch, M. A., Munk, W. H., and Rudnick, D. (**2004**), "Propagation of sound through a spicy ocean, the sofar overture," J. Acoust. Soc. Am. **116**(3), 1447–1462.

Franchi, E. R., and Jacobson, M. J. (**1972**), "Ray propagation in a channel with depth-variable sound speed and current," J. Acoust. Soc. Am. **52**(1), 316–331.

Flatte, S. M., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (1979), "Sound Transmission Through a fluctuating Ocean," (Cambridge University Press, Cambridge).

Fu, L. (**1981**), "Observations and models of inertial waves in the deep ocean," Rev. Geophys. Space Phys. **19**(1), 141–170.

Garrett, C. (**2001**), "What is the "near inertial" band and why is it different from the rest of the internal wave spectrum?," J. Phys. Oceanogr. **31**, 962–971.

Godin, O. A., and Voronovich, G. (**2004**), "Fermat's principle for non-dispersive waves in non-stationary media," J. Royal Soc. London **460**, 1631–1647.

Keller, J. B. (**1954**), "Geometric Acoustics. I: The theory of weak shock waves," J. Appl. Phys. **25**, 938–947.

Leaman, K. D., and Sanford, T. B. (**1975**), "Vertical propagation of inertial waves: A vector spectral analysis of velocity profiles," J. Geophys. Res. **80**, 1975–1978.

Lighthill, J. (**1978**), *Waves in Fluids*, Chap. 4 (Cambridge University Press, Cambridge, England).

Munk, W. (**1981**), "Internal waves and small scale processes," in *The Evolution of Physical Oceanography*, edited by C. Wunsch and B. Warren (MIT, Cambridge), 264–291.

Munk, W., Worcester, P., and Wunsch, C. (**1995**), *Ocean Acoustic Tomography*, Chap. 3 (Cambridge University Press, Cambridge, England).

Munk, W. H., Worcester, P. F., and Zachariasen, F. (**1981**), "Scattering of sound by internal wave currents: The relation to vertical momentum flux," J. Phys. Oceanogr. **11**, 442–454.

Pedlosky, J. (**1986**), *Geophysical Fluid Dynamics*, 2nd ed. (Springer-Verlag, Berlin).

Pierce, A. D. (**1989**), *Acoustics, An Introduction to its Physical Principles and Applications*, (Acoustical Society of America/AIP, Westbury).

Pinkel, R. (**1984**), "Doppler sonar observations of internal waves: The wavenumber frequency spectrum," J. Phys. Oceanogr. **14**, 1249–1270.

Robertson, J. S., Siegmann, W. L., and Jacobson, M. J. (**1985**), "Current and current shear effects in the parabolic approximation for underwater channels," J. Acoust. Soc. Am. **77**(5), 1768–1780.

Stallworth, L. A., and Jacobson, M. J. (**1972a**), "Sound transmission in an isospeed ocean channel with depth dependent current," J. Acoust. Soc. Am. **51**, 1738–1750.

Stallworth, L. A., and Jacobson, M. J. (**1972b**), "Acoustic propagation in a uniformly moving ocean channel with depth dependent sound speed," J. Acoust. Soc. Am. **52**, 344–355.

Stoughton, R. B., Flatté, S. M., and Howe, B. M. (**1986**), "Acoustic measurement of internal wave rms displacement and rms horizontal current off Bermuda in late 1983," J. Geophys. Res. **91**(6), 7721–7732.

Thompson, R. J. (**1972**), "Ray theory for an inhomogeneous moving medium," J. Acoust. Soc. Am. **51**, 1675–1682.

Uginčius, P. (**1965**), "Acoustic-ray equations for a moving inhomogeneous medium," J. Acoust. Soc. Am. **37**, 476–479.

Uginčius, P. (**1972**), "Fermats principle in a moving inhomogeneous medium," J. Acoust. Soc. Am. **51**, 1759–1763.

Worcester, P. F. (**1977**), "Reciprocal acoustic transmission in a midocean environment," J. Acoust. Soc. Am. **62**(4), 895–905.

Xu, J., Colosi, J. A., and Howe, B. M. (2005) (unpublished).

Confirmation of the Biot theory for water-saturated sands at high frequencies and effects of scattering on the attenuation of sound waves (L)

Keiichi Ohkawa[a)]
Division of Applied Marine Physics, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149

(Received 17 May 2005; revised 7 November 2005; accepted 10 November 2005)

Sound attenuation caused by the absorption and scattering of energy is studied. The Biot theory is used to predict the absorption coefficient. The scattering attenuation applies the experimental result. The calculated attenuation coefficient is the sum of absorption and scattering components, and is in excellent agreement with data collected during the sediment acoustics experiment in 1999. This implies that the frequency dependence of the attenuation coefficient due to the fluid viscosity follows $f^{1/2}$ in the high-frequency range, as the Biot theory predicts. This also suggests that, at high frequencies, the attenuation coefficient is not linear in f. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149770]

PACS number(s): 43.30.Ma, 43.30.Ky [RS] Pages: 709–711

I. INTRODUCTION

Half a century has passed since Biot presented the first paper regarding the propagation of elastic waves in a porous medium.[1,2] Biot originally intended to describe the propagation of elastic waves in a rock; however, his theory does not explain the dispersion of acoustic waves observed in rock. Although Biot assumes continuous pore spaces in a material, the pore fluid is actually isolated in cracks within the rock. Dvorkin and Nur introduced the squirt-flow mechanism[3] into the Biot theory, which successfully describes the wave dispersion.

After incorporating Stoll's large efforts on sediment acoustics,[4] the Biot theory predicts sound-wave speeds in water-saturated granular sediments rather well. But, the so-called Biot theory leads to an attenuation coefficient at low frequencies that is proportional to f^2 and at high frequencies that is proportional to $f^{1/2}$. Some data[5] suggest that the attenuation coefficient is proportional to $f^{1/2}$. Most data, however, suggest that the attenuation coefficient is linear in f over a wide range of frequencies.[6,7] More careful observation of data finds the attenuation coefficient includes both $f^{1/2}$ dependence in the middle-frequency range and dependence close to f in higher frequency ranges.[8] Nevertheless, most theories ignore the $f^{1/2}$ dependence of the attenuation coefficient to address the linear dependence in f. Another conflict is that most theories cannot simultaneously explain the frequency dependence both of sound speeds and attenuation coefficients. Recently, Buckingham[9] and Chotiros[10] separately developed new theories based on grain-to-grain shearing mechanisms. In both theories these mechanisms play an important role in high frequencies. However, effects of the shear are considered more important at low frequencies, which conflicts with intuitive notions.

Although scattering attenuation is generally ignored in the frequencies used in applications of underwater acoustics, effects of scattering are much bigger than predicted by a single scattering theory. Seifert *et al.* measured the scattering attenuation in water-saturated sand and found f^2 dependence in the attenuation coefficient in the ultrahigh frequency range.[11] The mean grain size used was 230 μm for fine sands. They compared the multiple forward scattering theory[12] with data and concluded that scattering is the dominant attenuation mechanism for unconsolidated sand with grain diameters of 230 μm or larger in the frequency range 100 to 1000 kHz.

In this paper, their result is used as the scattering attenuation, and the absorption coefficient is computed using the Biot theory.

II. COMPARISON OF THE THEORY AND DATA

The Biot theory predicts two kinds of longitudinal waves, whose wave numbers, l, for isotropic materials are given by solving the secular equation

$$\begin{vmatrix} Hl^2-\rho\omega^2 & Cl^2-\rho_f\omega^2 \\ Cl^2-\rho_f\omega^2 & Ml^2-m'\omega^2 \end{vmatrix} = \mathbf{0}, \tag{1}$$

with $m'=m-i(\eta/\omega k_0)\sqrt{1+i(N\omega/2\omega_c)}$, where ρ is the bulk density of the porous medium, ρ_f the fluid density, ω the angular frequency, ω_c the characteristic frequency, m the added mass density, k_0 the dc permeability, and η the fluid viscosity. Elastic moduli H, C, and M are expressed in terms of bulk and shear moduli and the porosity of the porous material. The modulus N is the constant determined by the structure of the medium using the concept of the dynamic tortuosity and permeability.[13] For most porous media, $N\approx 1$ can be assumed.[14]

The absorption coefficient α_a (dB/m) is given by

$$\alpha_a = 20\log_{10} e\cdot \mathrm{Im}(l_1), \tag{2}$$

where l_1 is the phase velocity of Biot's fast wave.

[a)]Presently at the 5th Research Center, Technical R&D Center, Japan Defense Agency, 3-13-1 Nagase Yokosuka, 2390826, Japan. Electronic mail: ohkawa@jda-trdi.go.jp

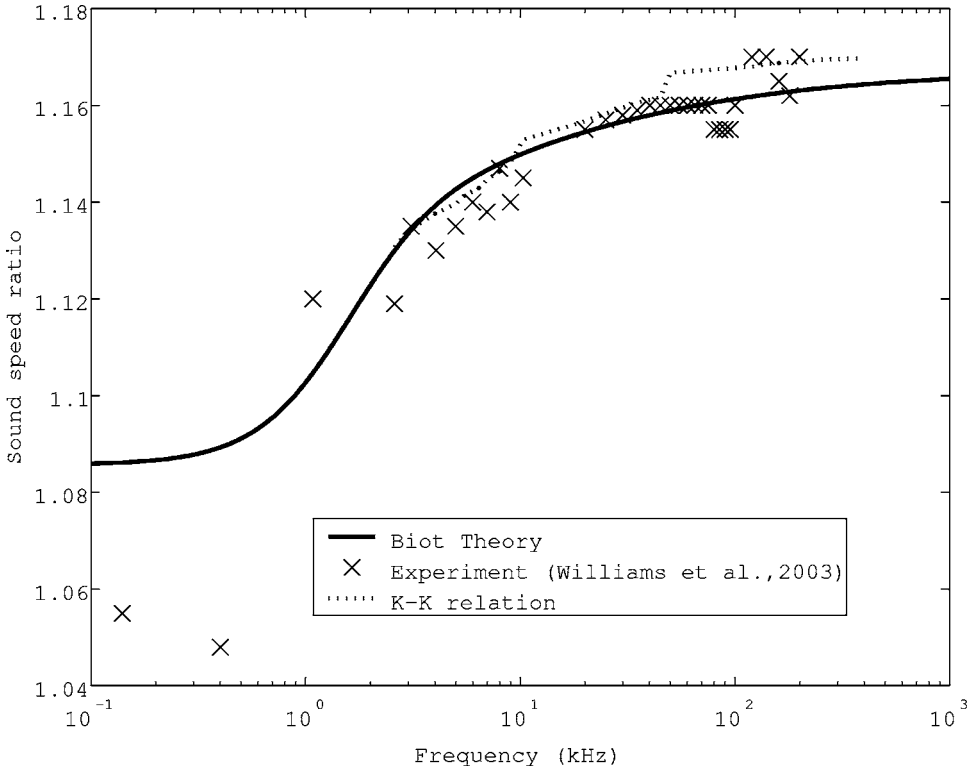


FIG. 1. Comparison of sound-speed ratio and data.

From the experimental result for water-saturated fine sands,[11] the scattering attenuation α_s (dB/m) as a function of frequency fits

$$\alpha_s = 1.2 \times 10^{-10} \cdot 20 \log_{10} e \cdot f^2, \tag{3}$$

where the units of f is hertz.

The algebraic sum of the above components gives the total attenuation coefficient,

$$\alpha = \alpha_a + \alpha_s. \tag{4}$$

The phase velocity and attenuation coefficient are compared with the experimental data collected during the sediment acoustics experiment in 1999 (SAX 99).[8,15] The model parameters are the same as those in Williams *et al.*[8] Figure 1 shows the sound-speed ratio is in excellent agreement with the experimental result. Figure 2 shows the calculated attenu-

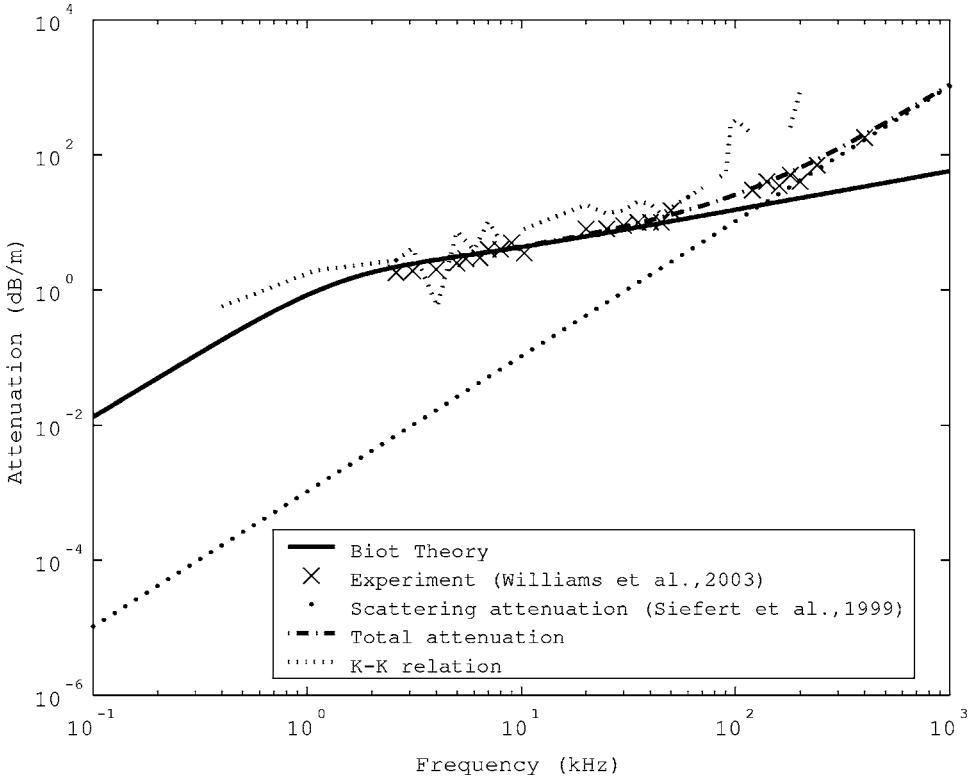


FIG. 2. Comparison of attenuation coefficient and data.

ation coefficient is also in excellent agreement with the data for overall frequencies. In Fig. 2, each component of the attenuation coefficient is depicted. As seen from the figure, intrinsic attenuation is dominant at frequencies below 50 kHz, while scattering is dominant at frequencies greater than 200 kHz. The mean grain size of sand at the site of SAX 99 ranges from 400 to 500 μm, so the actual scattering attenuation may be larger. The frequency dependence of sound-wave attenuation in water-saturated sands is quite similar to that of sound absorption in seawater.

Finally, the sound-speed dispersion and absorption coefficient are recalculated via the Kramers-Krönig (K-K) relationship[16–18] and replotted on the sound-speed and attenuation data. The sound-speed dispersion (dashed lines in Fig. 1) is calculated using the intrinsic attenuation, which is estimated by the measured attenuation−Eq. (3). The dispersion curve is found to follows the Biot theory. In turn, the absorption coefficient is calculated using the measured sound speed and plotted in Fig. 2. The absorption coefficient (dashed lines) follows the Biot theory in the high-frequency range from 1 to 50 kHz. The large deviation from the data at higher frequencies may be due to the measurement error.

III. CONCLUSION

Effects of scattering on the attenuation are much greater than expected in the frequency range used in underwater acoustics. The calculated attenuation coefficient is in excellent agreement with data collected during the SAX 99 experiments for overall frequencies. This suggests that the frequency dependence of the attenuation coefficient due to fluid viscosity follows $f^{1/2}$ in the high-frequency range, as the Biot theory predicts. This frequency dependence is confirmed via the Kramers-Kronig relationship. The frequency dependence of the attenuation coefficient was not found to be linear in f in the data at high frequencies. The sound speed measured at high frequencies is also in excellent agreement not only with the Biot theory but also with the K-K relationship. Therefore, it is concluded that the Biot theory predicts the sound speed and intrinsic attenuation in water-saturated sands at high frequencies and that the deviation of attenuation from the Biot theory is due to effects of scattering.

[1]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[2]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[3]J. Dvorkin and A. Nur, "Dynamic poroelasticity: A unified model with the squirt and the Biot mechanisms," Geophysics **58**, 524–533 (1993).
[4]D. R. Stoll, *Sediment Acoustics* (Springer, New York, 1989).
[5]J. M. Hovem and G. D. Ingram, "Viscous attenuation of sound in saturated sand," J. Acoust. Soc. Am. **66**, 1807–1812 (1979).
[6]E. L. Hamilton and R. T. Bachman, "Sound velocity and related properties of marine sediments," J. Acoust. Soc. Am. **72**, 1891–1904 (1982).
[7]D. J. Wingham, "The dispersion of sound in sediment," J. Acoust. Soc. Am. **78**, 1757–1760 (1985).
[8]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).
[9]M. J. Buckingham, "Theory of acoustic attenuation, dispersion, and pulse propagation in unconsolidated granular materials including marine sediments," J. Acoust. Soc. Am. **102**, 2579–2596 (1998).
[10]N. P. Chotiros and M. Isakson, "A broadband model of sandy ocean sediments: Biot–Stoll with contact squirt flow and shear drag," J. Acoust. Soc. Am. **116**(4), 2011–2022 (2004).
[11]P. K. Seifert, B. Kaelin, and L. R. Johnson, "Effect on ultrasonic signals of viscous pore fluids in unconsolidated sand," J. Acoust. Soc. Am. **106**, 3089–3094 (1999).
[12]B. Kaelin and L. R. Johnson, "Dynamic composite elastic medium theory. II. Three-dimensional media," J. Appl. Phys. **84**, 5458–5468 (1998).
[13]L. D. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[14]N. D. Albert, "A comparison between wave propagation in water-saturated and air-saturated porous materials," J. Appl. Phys. **73**(1), 28–36 (1993).
[15]M. D. Richardson *et al.*, "Overview of SAX99: Environmental Considerations," IEEE J. Ocean. Eng. **26**, 26–53 (2001).
[16]R. Kronig, "On the theory of dispersion of x-rays," J. Opt. Soc. Am. **12**, 547–557 (1926).
[17]R. Kronig and H. A. Kramers, "Absorption and dispersion in x-ray spectra," Z. Phys. **48**, 174 (1928).
[18]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**(3), 695–701 (1981).

Comment on "The cochlear amplifier as a standing wave: 'Squirting' waves between rows of outer hair cells?" [J. Acoust. Soc. Am. 116, 1016–1024 (2004)] (L)

Eric L. LePage[a)]
OAEricle Laboratory, P. O. Box 6025 Narraweena, NSW, 2099 Australia

(Received 15 February 2005; revised 4 November 2005; accepted 7 November 2005)

Bell and Fletcher [J. Acoust. Soc. Am. **116**, 1016–1024 (2004)] proposed that one of the functions of activity of the outer hair cells (OHCs) might be a fluid-pumping action generating lateral fluid flow in the gap between the reticular membrane and the tectorial membrane and they supplied mathematical and descriptive justification for their theory which drew heavily upon the postulation (Gold, 1948) of the need for an active mechanism in the mammalian cochlea. In the 1970s there had been considerable speculation about how the inner hair cell (IHC) stereocilia are stimulated, whether they are stimulated in proportion to basilar membrane displacement or velocity or both, and whether the velocity dependence is due to subtectorial fluid flow. In 1977 experiments were conducted to investigate the possibility of subtectorial fluid flows using a dye as tracer. The work was not reported because it had been conducted at a time when visual observation of cochlear function had fallen out of favor in comparison with the more sensitive techniques thought necessary to observe submicroscopic phenomena, and secondly because it yielded a negative result. The essential details of those experiments are reported here to note for the record the extent to which this elaborate idea has already been tested. [DOI: 10.1121/1.2146087]

PACS number(s): 43.64.Bt, 43.64.Kc [BLM] Pages: 712–714

I. INTRODUCTION

In the 1960s and 1970s the predominant interest in fluid motion stemmed from the various descriptions of eddies by Békésy (1960, p. 420) in connection with the generation of a traveling wave along the cochlear partition. His greatest impact arose from his visual observations using stroboscopic illumination leading to his traveling-wave theory. The conclusion was that the displacements of the vibrating structures in cadaver ears were exceedingly small at low sound levels. As a result, direct cochlear mechanical measurements have since employed increasingly sophisticated techniques with sensitivity thought necessary to reveal any useful result.

At the time, the traveling wave theory led to extensive fluid mechanical modeling [e.g., by Lesser and Berkeley (1972)] where again the emphasis was on longitudinal movement in scala vestibuli and scala tympani. The notion of fluid movements within the organ of Corti, either in the tunnel below the reticular membrane, or in the subtectorial space, at the time, continued to be discounted because of fluid viscosity. The proposal by Gold (1948) that the cochlea should contain an active mechanism to overcome viscous effects drew some attention at the time, yet his ideas were still largely ignored in the 1970s despite new findings of cochlear nonlinearity (Rhode, 1971). Models of oscillatory subtectorial fluid flow were certainly published (Billone and Raynor, 1973; Zwislocki and Sokolich, 1973; Crandall, 1975) and subsequently were addressed by the elegant morphological studies of Lim (1986), but still they seemed remote because of reservations about fluid damping the vibration of the structures preventing any resonance, primarily due to boundary layer and fluid viscosity considerations. Of interest was a model which suggested that instead of the flow being oscillatory, the major function may be a streaming flow across the subtectorial gap (Steele, 1973). This possibility continues to generate interest because of the need to account for the existence and rectification inherent in the summating potential, which changes polarity along the traveling wave envelope (Davis, *et al.*, 1958; Dallos *et al.*, 1972). Such rectification may occur not just because of hair cell transducer characteristics, but also because the possibility that cochlear activity leads to a mechanical correlate of cochlear displacement-baseline regulation (LePage, 1987). Such an "offset" may have as its origin the motility of the OHC or indeed streaming fluid flow in the subtectorial space.

Since the revelation of two types of OHC activity [summary (Kros, 2004)] there has been a revival of interest in the notion of fluid flows, including the reports concerning flow in the organ of Corti (Karavitaki and Mountain, 2003) and in the gap between the reticular membrane and the tectorial membrane (Bell and Fletcher, 2004). The latter authors postulated that OHC activity may result in some pumping action producing subtectorial fluid flow.

The purpose of this letter is to report one experiment which was conducted specifically to investigate the possibility of subtectorial fluid movement. These studies were carried out in a number of preparations which took place over a period from 27 June to 27 July 1977. The results were not reported at the time, because, firstly, with the realization that Békésy had not seen any motions of interest except at very high sound levels, visual observations had fallen out of favor and been replaced by more sensitive techniques. Secondly,

[a)]Electronic mail: ericlepage@oaericle.com.au

they were not published because the simple experiment of instilling tracer dye to look for fluid movement produced a negative result. The experiments were conducted in the same series which confirmed Rhode's report of nonlinear mechanical activity (LePage and Johnstone, 1980), so, while the actual thresholds were not recorded, some tuning-related flows should have been observed. It is clear that the approach is worthy of repetition with the latest techniques.

II. METHODS

Pigmented guinea pigs of 200–250 g were prepared according to standard procedures (LePage and Johnstone, 1980) and ethical guidelines. Methylene blue in artificial endolymph solution (150 mM potassium chloride) was the dye selected because it can be detected visually even in very high levels of dilution. A small hole was shaved in the outer bony wall of the cochlea, above the basilar membrane. The tip of the injecting glass pipette had a diameter of 2 to 5 μm and was therefore smaller than a hair cell. The procedure was visualized through a Zeiss OP1 operating microscope with its internal forward illumination. Backward fiber optic illumination through the whole cochlea was tried but was less satisfactory. Microliter quantities of the dye were injected into each of the fluid spaces considered, at a place about 3.5 mm from the stapes. The results for each of the chambers was visually distinguishable. In preliminary experiments, if the dye was injected into any of the scala, tympani, scala vestibuli, or scala media a diffuse distribution of dye occurred. The aim was to deliver dye precisely into scala media so that any possible toxicity of the dye solution which might otherwise occur (Zenner *et al.*, 1994) could be discounted. If the tip of the electrode was located in either of the two longitudinal channels—the outer and inner sulci—the dye flowed along the bony channel and the edges of the deep blue region were highly defined. The lower resistance to flow of the dye along these channels meant that the experiment could be performed away from the exact site of injection, limiting possible injection artifacts. Once the micro-bolus of dye was instilled, the pipette was removed.

Sound stimuli were, for the major part, pure tones. These were delivered by an electrostatic transducer (B&K4134 biased with 200 V dc voltage) using techniques described in LePage (1987). The tones were varied in frequency and level appropriate for the place under view over their physiological ranges (i.e., 10 to 25 kHz, 20 to 80 dB SPL) covering conditions required to see sharp tuning.

III. RESULTS

The results of these experiments were typically obtained within an hour of beginning the preparation. The hearing thresholds were routinely monitored using the N_1 whole nerve action potential threshold (LePage and Johnstone, 1980) and low thresholds could be expected for up to one hour after that time. An important variable turned out to be the pressure of the dye injection. Less pressure was required to instill small localized deposits of dye into the inner or outer sulcus, whereupon the dye would progress longitudinally along either sulcus. This provided visual feedback that the targeted cavities were found. With high pressure injection, it was possible to see localized radial movement of the dye in the subtectorial space under pressure alone at the site of injection. However, in general, *no stimulus-controlled dye movement or accelerated dilution was visible*. Various other stimuli, continuous and pulsed, were tried. The results compelled us to reject the hypothesis that subtectorial fluid flow was the method by which the OHC delivered their output to the IHC stereocilia.

IV. DISCUSSION

If Bell and Fletcher's "squirting" theory could be verified, there should have been clear evidence in our 1977 experiments that the movement of the dye should have been gated by starting and stopping sound for at least one frequency in the range tried. By implication this would mean that OHC motility is indeed efficient enough to overcome viscous boundary layer effects (Scherer and Gummer, 2004).

Of particular interest is whether Bell and Fletcher's term "squirting" implies purely oscillatory fluid motion or also entails some streaming fluid motion (net fluid transport presumably towards the IHC stereocilia). However, even if only active squirting in alternate directions with each half cycle is envisaged, some accelerated diffusion of the dye at the edges of the subtectorial region should still have been visible.

Why did such a simple experiment fail to show dye movement or dilution? A common perception is that the condition of guinea-pig preparations carried out in the 1970s was not adequate to show cochlear activity and that the first good measurements only took place a decade later (Sellick, *et al.*, 1982). The connection between sharply tuned nonlinear behavior and hearing threshold was established three years earlier, because LePage and Johnstone (1980) routinely recorded animal N_1 action potential threshold. Those experiments were in fact the ones which first established the decline in the degree of nonlinear compression with decline in sensitivity. Moreover, those same experiments clearly demonstrated the existence of sharp tuning in the frequency dependence of the nonlinear gradient (LePage, 1981) and pointed to the need to go to lower sound levels. Those early experimental data showed categorically that it was not necessary to see sharp tuning as a conventionally plotted tuning curve, in order to see this sharply tuned effect in the frequency dependence of the input-output curve. The relevance here is that it is unlikely that the dye-tracer experiments reported here to look for activity-driven fluid pumping were conducted on preparations devoid of OHC activity.

Secondly, it is possible that the dye would not be visible at dilutions which might have occurred corresponding to threshold sound stimulation. It is felt this possibility is discounted because the Methylene blue dye is intense for the concentration used and was still discernable in dilutions of three orders of magnitude. This led to our expectation of observable streaming effects for stimulations relevant to the operation of the cochlear amplifier, i.e., for stimulations of below 50 dB SPL.

Thirdly, in keeping with the Gold hypothesis, the resonance may be very high Q ($\sim$100). The most basic resonant

element may therefore be tuned so highly that we scanned too rapidly through the most appropriate frequency and level to see any streaming ("squirting" action). This possibility now seems unlikely. This is because at places in the basal turn, the region displaying neural and mechanical resonance has a $Q_{10\ dB}$ of only 5–10; the equivalent distance span along the cochlear partition is approximately 500 μm. If the sharp tip of the tuning curve is made up of many even-more-sharply stagger-tuned processes in parallel, we should still have seen migration of the dye through bulk flow or accelerated dispersion of the dye due to resonance. Moreover, if streaming flow were to occur, it should not only have been strongly place dependent, but with flow reversal should have been visible at some place, because fluid streaming in a closed vessel necessitates a circuit and most likely eddies as well.

Fourthly, Bell and Fletcher's Fig. 1 is drawn assuming that the subtectorial gap extends out to the open scala media which is not consistent with the descriptions of Lim (1986), which depict connection of the outer edge of the tectorial membrane to the Hensen cells via the marginal net. It is possible that this attachment of the tectorial membrane to the Hensen cells may have provided partial rectification of the flow (allowing only outward movement), thus preventing the dye from moving from the outer sulcus into the subtectorial gap. However, those experiments which actually did see dye within the subtectorial gap in the region of the OHC originating in the inner sulcus nevertheless showed no stimulus controlled trace of dye.

As observed above, these experiments do not test the notion of important fluid movement in the region of the cell bodies below the reticular membrane. In this respect Bell and Fletcher's title is somewhat ambiguous, because it is too general for the main point of their theory. Their article is mainly directed at fluid movement above the reticular membrane such as was directly tested in these experiments. Their theory nevertheless allows that some lateral fluid movement may also take place between the cell bodies below the reticular membrane, but it is not at all clear whether such a possibility has any relevance to stimulation of the IHC stereocilia.

In the current era of very detailed and cochlear mechanical measurements the dye tracer experiment still stands out as a valid approach and should not be subject to the criticisms of Békésy's visual observations. It was not looking for motions of microscopic structures so much as evidence of fluid pumping action due to OHC activity. This letter points to the need for a new experiment in a preparation with sharp tuning demonstrated with modern techniques which include recording tracer movement simultaneously, preferably using a method that can distinguish between somatic and hair bundle motility.

ACKNOWLEDGMENTS

I thank Dr. Brian M. Johnstone for the formative role he played during this period.

Békésy, G. v. (**1960**). *Experiments in Hearing*. (McGraw-Hill, New York).
Bell, A., and Fletcher, N. H. (**2004**). "The cochlear amplifier as a standing wave: 'squirting' waves between rows of outer hair cells?" J. Acoust. Soc. Am. **116**, 1016–1024.
Billone, M., and Raynor, S. (**1973**). "Transmission of radial shear forces to cochlear hair cells," J. Acoust. Soc. Am. **54**, 1143–1156.
Crandall, W. E. (**1975**). "Diode pump cochlear audition theory," Int. J. Neurosci. **6**(5), 203–213.
Dallos, P., Schoeny, Z. G., and Cheatham, M. A. (**1972**). "Cochlear summating potentials. Descriptive aspects," Acta Oto-Laryngol., Suppl. **302**, 1–46.
Davis, H., Deatherage, B. H., Eldredge, D. H., and Smith, C. A. (**1958**). "Summating potentials of the cochlea," Am. J. Physiol. **195**, 251–261.
Gold, T. (**1948**). Hearing II. "The physical basis of the cochlea,", Proc. R. Soc. London (B)., **135**, 492–498
Karavitaki, K. D. , and Mountain, D. C. (**2003**). "Is the cochlear amplifier a fluid pump?" in Biophysics of the Cochlea: from Molecules to Models. Edited: A. W. Gummer (World Scientific Publishing, Singapore ISBN: 981–238–304–2), 310–311.
Kros, C. (**2005**). "Hearing: aid from hair force," Nature (London) **433**(7028), 810–811.
LePage, E. L. (**1981**). "The role of nonlinear mechanical processes in mammalian hearing," unpublished doctoral dissertation, The University of Western Australia.
LePage, E. L. (**1987**). "Frequency-dependent self-induced bias of the basilar membrane and its potential for controlling sensitivity and tuning in the mammalian cochlea," J. Acoust. Soc. Am. **82**, 139–154.
LePage, E. L., and Johnstone, B. M. (**1980**). "Nonlinear mechanical behavior of the basilar membrane in the basal turn of the guinea pig cochlea," Hear. Res. **2**, 183–189.
Lesser, M. B., and Berkley, D. A. (**1972**). "Fluid mechanics of the cochlea. Part 1," J. Fluid Mech. **51**, 497–512.
Lim, D. J. (**1986**). "Functional structure of the organ of Corti: a review," Hear. Res. **22**, 117–146.
Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique," J. Acoust. Soc. Am. **49**, Suppl. 2, 1218.
Scherer, M. P., and Gummer, A. W. (**2004**). "Vibration pattern of the organ of Corti up to 50 kHz: evidence for resonant electromechanical force," Proc. Natl. Acad. Sci. U.S.A. **101**(51), 17652–17657.
Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). "Measurement of basilar membrane motion in the guinea pig using the Mossbauer technique," J. Acoust. Soc. Am. **72**, 131–141.
Steele, C. R. (**1973**). "A Possibility for Sub-Tectorial Membrane Fluid Motion," in *Basic Mechanisms in Hearing* (Academic, London), pp. 69–93.
Zenner, H. P., Reuter, G., Zimmermann, U., Gitter, A. H., Fermin, C., and LePage, E. L. (**1994**). "Transitory endolymph leakage induced hearing loss and tinnitus depolarization, biphasic shortening and loss of electromotility of outer hair cells," Eur. Arch. Otorhinolaryngol. **251**(3), 143–153.
Zwislocki, J. J., and Sokolich, W. G. (**1973**). "Velocity and displacement responses in auditory-nerve fibers," Science **182**(107), 64–66.

Technique for "tuning" vocal tract area functions based on acoustic sensitivity functions (L)

Brad H. Story[a)]
Speech Acoustics Laboratory, Department of Speech and Hearing Sciences, University of Arizona, Tucson, Arizona 85721

(Received 17 May 2005; revised 17 October 2005; accepted 22 November 2005)

A technique for modifying vocal tract area functions is developed by using sum and difference combinations of acoustic sensitivity functions to perturb an initial vocal tract configuration. First, sensitivity functions [e.g., Fant and Pauli, Proc. Speech Comm. Sem. **74**, 1975] are calculated for a given area function, at its specific formant frequencies. The sensitivity functions are then multiplied by scaling coefficients that are determined from the difference between a desired set of formant frequencies and those supported by the current area function. The scaled sensitivity functions are then summed together to generate a perturbation of the area function. This produces a new area function whose associated formant frequencies are closer to the desired values than the previous one. This process is repeated iteratively until the coefficients are equal to zero or are below a threshold value. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151802]

PACS number(s): 43.70-h, 43.70.Bk [AL] Pages: 715–718

I. INTRODUCTION

The shape of the vocal tract can be approximately represented by an area function; that is, the variation in cross-sectional area as a function of distance from the glottis. A pattern of acoustic resonances can be calculated based on the shape of any given area function, and will indicate the locations of the formant frequencies that contribute to both phonetic and speaker-specific characteristics. It may be of interest to know how particular changes in the formant frequency pattern could be generated by changes to the shape of the area function, and vice versa. For example, how might an area function for the vowel [ɑ] be modified so that the second formant (F2) is increased in frequency while all other formants remain fixed at their original values? Or perhaps the interest may be in altering the area function to generate a particular pattern of the upper formants (e.g., F3–F5) such that a distinct change in sound quality is produced, while F1 and F2 remain fixed.

The purpose of this letter is to present a technique that iteratively adjusts ("tunes") the shape of a given vocal tract area function so that a specific pattern of formant frequencies is produced. The technique is based on perturbing the shape of an initial vocal tract configuration with a summation of scaled acoustic sensitivity functions, such that the formants are systematically displaced toward desired values. Although it is well known that transformation of a set of formant frequencies to a vocal tract area function does not produce a unique solution (e.g., Schroeder, 1967; Mermelstein, 1967; Wakita, 1973; Atal, Chang, Mathews, and Tukey, 1978; Milenkovic, 1984; Sondhi and Resnick, 1983; Sorokin, 1992), the technique described may be useful for generating subtle, hypothetical, modifications to a specific vocal tract shape.

[a)]Electronic mail: bstory@u.arizona.edu

II. ACOUSTIC SENSITIVITY FUNCTIONS

The sensitivity of a particular formant frequency to a change in vocal tract cross-sectional area can be defined as the difference between the kinetic energy (KE) and potential energy (PE) as a function of distance from the glottis, divided by the total energy in the system (Fant and Pauli, 1975). A sensitivity function can be written as

$$S_n(i) = \frac{\mathrm{KE}_n(i) - \mathrm{PE}_n(i)}{\mathrm{TE}_n} \quad n = 1,2,3, \ldots \quad \text{and}$$
$$i = [1, \ldots, N_{\text{areas}}], \tag{1}$$

where i is the section number (section 1 is just above the glottis and section N_{areas} is at the lips), n is the formant number, and

$$\mathrm{TE}_n = \sum_{i=1}^{N_{\text{areas}}} [\mathrm{KE}_n(i) + \mathrm{PE}_n(i)]. \tag{2}$$

The kinetic and potential energies for each formant frequency are based on the pressure $P_n(i)$ and volume velocity $U_n(i)$ computed for each section of an area function. They are calculated as

$$\mathrm{KE}_n(i) = \frac{1}{2}\frac{\rho l(i)}{a(i)}|U_n(i)|^2 \tag{3}$$

and

$$\mathrm{PE}_n(i) = \frac{1}{2}\frac{a(i)l(i)}{\rho c^2}|P_n(i)|^2, \tag{4}$$

where $a(i)$ and $l(i)$ are the cross-sectional area and length of element i within an area function, respectively, and ρ is the density of air and c is the speed of sound. The area functions used throughout this study contained 44 sections, each with a length of $l(i)=0.396825$ cm. Hence $i=[1,\ldots,44]$ and the actual distance from the glottis for each section is $x = i \cdot l(i)$.

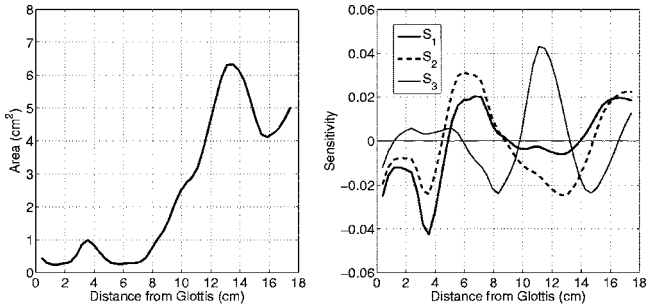


FIG. 1. Sensitivity function calculation for a male [ɑ] vowel based on Story, Titze, and Hoffman (1996). (a) Measured area function, (b) sensitivity functions for F1, F2, and F3.

Calculations of pressures, flows, and frequency response functions for this study were accomplished with a transmission-line type model of the vocal tract (e.g., Sondhi and Schroeter, 1987; Story, Laukkanen, and Titze, 2000) that included energy losses due to yielding walls, viscosity, heat conduction, and acoustic radiation at the lips. This particular implementation did not, however, include any side branches such as the piriform sinuses, sublingual cavities, or nasal passages. While it is recognized that these cavities may significantly affect some formant frequencies (Dang and Honda, 1997; Espy-Wilson, 1992; Makarov and Sorokin, 2004), their omission here does not affect the development of the proposed method, nor would it prevent their inclusion in the future.

As an example, sensitivity functions were calculated for a male [ɑ] vowel based on Story, Titze, and Hoffman (1996) [see Fig. 1(a)] and are shown in Fig. 1(b). Each line extends along the distance from the glottis to lips and indicates the relative sensitivity of the first, second, and third formants (F1, F2, and F3) to a small perturbation of the area function $[\Delta a(i)]$. Mathematically, this can be written as,

$$\frac{\Delta \mathrm{F}_n}{\mathrm{F}_n} = \sum_{i=1}^{N_{\text{areas}}} S_n(i) \frac{\Delta a(i)}{a(i)}, \tag{5}$$

where n is again the formant number. Using S_1 in Fig. 1(b) and Eq. (5) as a guide, it is observed that F1 could be increased by expanding the area in regions along the vocal tract length between 5 cm and 9 cm from the glottis and from 14 cm to the lip termination. F1 could also be increased by constricting the regions between the glottis and 5 cm, as well as between 9 cm and 14 cm. Lowering F1 would require the opposite changes in area within the same regions. For S_2, an increase in F2 could be produced by expanding the regions between 4.5–8.8 cm and 14.8–17.5 cm, and constricting the regions of the area function that extend from 0 to 4.5 cm and 8.8 to 14.8 cm; lowering F2 would require the opposite changes in area. Changes in F3 could be similarly carried out by modifying cross-sectional areas in the positively and negatively valued regions specified by S_3. Although not shown, sensitivity functions corresponding of F4 and F5 were also calculated.

III. AREA FUNCTION PERTURBATION

Whereas changes to an area function that would modify formant frequencies according to the calculated sensitivity functions can be performed manually (e.g., Story, Titze, and Hoffman, 2001), an automated process would be more efficient, and ultimately more useful. The proposed technique consists of superimposing scaled (to affect cross-sectional area) replicas of the sensitivity functions on an area function. Direct superposition of S_1 for any area function would raise F1, whereas its opposite, $-S_1$, would lower it. F2 could be similarly controlled with superposition of a scaled S_2 replica, where $+S_2$ would increase F2 and $-S_2$ would decrease it. Higher frequency formants (e.g., F3, F4, and F5) could also be controlled with superposition of their respective sensitivity functions. Shifting multiple formants simultaneously could be carried out with superposition of the sum of $\pm S_1, \pm S_2, \pm S_3, \ldots, \pm S_n$.

The prediction of formant frequency change based on sensitivity functions is, however, limited to small area changes (approximately $<10\%$). Thus, sensitivity functions need to be recomputed after any small amount of area change, and a new perturbation determined. This can be performed iteratively until arriving at an area function that produces a desired set of formant frequencies. The process is mathematically represented as,

$$a_{k+1}(i) = a_k(i) + \sum_{n=1}^{N_{\text{fmts}}} z_{n_k} S_{n_k}(i)$$
$$i = [1, N_{\text{areas}}] \quad k = [0, N_{\text{iter}}]. \tag{6}$$

With the initial area function denoted by $a_0(i)$, the $a_k(i)$'s and S_{n_k}'s are vocal tract area functions and sensitivity functions, respectively, at successive iterations. The coefficients z_{n_k} scale the sensitivity functions so the area function perturbation displaces the formant frequencies in the desired direction. At every iteration, the z_{n_k}'s are determined by

$$z_{n_k} = \alpha \left[\frac{\mathcal{F}_n - \mathrm{F}_{n_k}}{\mathrm{F}_{n_k}} \right], \tag{7}$$

where $\mathcal{F}_n$ is a set of target formant frequencies and the F_{n_k}'s are the formants that correspond to the kth area function. α is an additional scale factor that can be used to speed the iterative process and is typically set to $\alpha = 10$. The iterations continue until the root of the sum of the squared differences between target formants and those of the kth area function,

$$\delta = \sqrt{\sum (\mathcal{F}_n - \mathrm{F}_{n_k})^2} \tag{8}$$

is less than a desired tolerance value. For this study, the iterations were allowed to proceed until $\delta < 0.1$ Hz.

To protect against cross-sectional areas becoming too small over the course of successive iterations, the superposition is performed logarithmically for those areas within an area function that are less than 1 cm^2 so that Eq. (6) becomes,

$$a_{k+1}(i) = \begin{cases} a_k(i) + \sum_{n=1}^{N_{\text{fmts}}} z_{n_k} S_{n_k}(i) & \text{for } a_k(i) > 1 \\ \exp\left(\ln(a_k(i)) + \ln\left(\sum_{n=1}^{N_{\text{fmts}}} z_{n_k} S_{n_k}(i) + 1 \right) \right) & \\ \quad \text{for } a_k(i) \leqslant 1. & \end{cases} \tag{9}$$

In addition, a minimum area threshold is set such that,

$$a_{k+1}(i) = \max[a_{k+1}(i), 0.1], \tag{10}$$

where the 0.1 is in units of square centimeters.

It is noted that Carré (2004) proposed a similar iterative technique for modifying the shape of an area function, specifically a uniform tube. His technique was based on using sensitivity functions as deformation patterns, but did not include individual scaling coefficients like the z_{n_k}'s specified here. Instead, each sensitivity function was always scaled with an amplitude of ±1.0, as well as a "deformability function" that constrained specific regions of the area function. Because of the constraints on the scaling coefficients, Carré's method apparently does not allow for the specification of a desired formant pattern, but rather modifies the area function successively with constant coefficients to produce a varying formant contour.

IV. EXAMPLES OF AREA FUNCTION TUNING

To demonstrate the method outlined in the previous section, the area function shown in Fig. 1(a) ([ɑ] vowel) was "tuned" in two different ways. In the first example, the second formant was shifted upward in frequency while all other formants (F1, F3, F4, and F5) were held constant at their original calculated values. For the second example, the upper formants were shifted toward each other so that they formed a cluster in the vicinity of about 3000 Hz [this is a typical frequency range of the "singing formant" (Sundberg, 1974)].

A. Upward shift of F2

With the [ɑ] area function (Fig. 1) serving as the initial vocal tract configuration $a_0(i)$, the corresponding initial (calculated) formant frequencies were,

$$\mathrm{F}_{n_0} = 800, 1136, 2770, 3448, 4240 \text{ Hz}, \quad n = 1,2,3,4,5 \tag{11}$$

and the target set of formant frequencies were set to,

$$\mathcal{F}_n = 800, \mathbf{1400}, 2770, 3448, 4240 \text{ Hz} \quad n = 1,2,3,4,5, \tag{12}$$

where only F2 differs from the original formants. Based on Eq. (7), and with α=10, the initial z coefficients were

$$z_{1_0} = 0.0, \tag{13}$$

$$z_{2_0} = 10\left[\frac{\mathcal{F}_2 - F_{2_0}}{F_{2_0}} \right] = \left[\frac{1400 - 1136}{1136} \right] = 2.32, \tag{14}$$

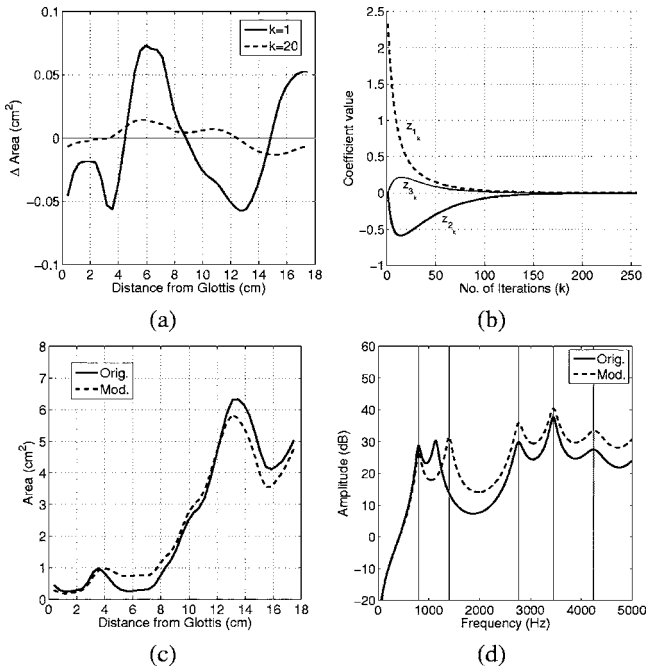


FIG. 2. Example of area function tuning to produce an upward shift of F2 based on the [ɑ] vowel in Fig. 1(a). (a) Area function perturbations at the first (solid) and the twentieth (dashed) iterations. (b) Values of the z for the first three coefficients as they converge toward 0.0 over 257 iterations. (c) Initial (solid) and modified (dashed) area functions. (d) Frequency response functions of the initial (solid) and final (dashed) area functions; the vertical lines represent the target formant frequencies.

$$z_{3_0} = z_{4_0} = z_{5_0} = 0.0. \tag{15}$$

The perturbation pattern imposed on the initial area function at the first iteration is simply $2.32S_2$ and is plotted as the solid line in Fig. 2(a). Note that the y-axis is shown as "Δ area cm^2" to indicate the cross-sectional area that is added to, or subtracted from, the area function and represents $\Delta a(i)$ in Eq. (5). The shape of the perturbation will gradually evolve over the course of the iterations to reflect the progressive change in sensitivity of the modified area function. As an example, the perturbation at the 20th iteration is shown as the dashed line.

After 257 iterations, $\delta < 0.1$ Hz and the z coefficients corresponding to each formant have converged toward zero as shown for the first three formants in Fig. 2(b); the z coefficients for F4 and F5 similarly converge but are not shown in order to preserve the clarity of the figure. The original and final area functions are shown in Fig. 2(c), where it can be seen that the tuning process has generated an expansion of the pharyngeal part of the vocal tract, and a reduction of the cross-sectional areas in the oral cavity and just above the glottis. The corresponding frequency response functions are plotted in Fig. 2(d) and indicate that the second formant frequency produced by the modified area function is precisely matched to the target frequency of 1400 Hz. The other formants were successfully maintained at their original values.

B. Clustering of F3, F4, and F5

A second example of area function tuning is shown in Fig. 3. The initial area function was the same [ɑ] vowel used in the first example. As before, the initial calculated formants were,

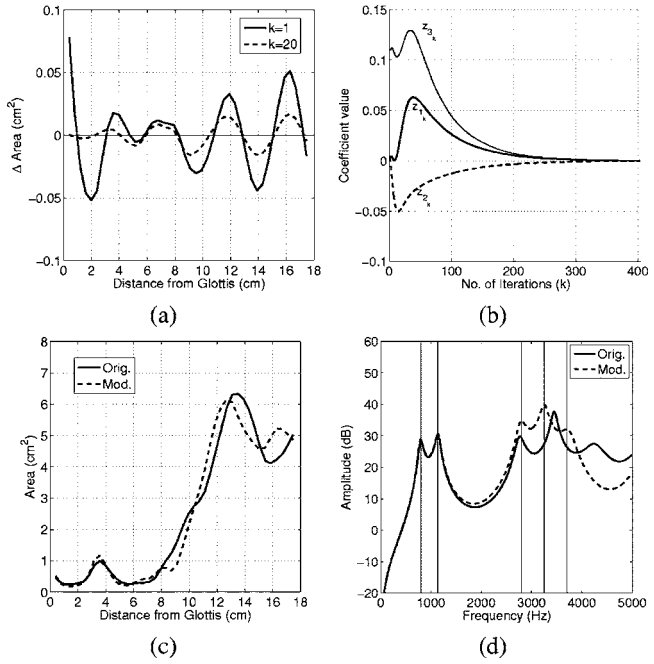


FIG. 3. Example of area function tuning to compress the distance between F3, F4, and F5, while maintaining F1 and F2 at their original values. (a) Area function perturbations at the first (solid) and the twentieth (dashed) iterations. (b) Values of the z coefficients as they converge toward 0.0 over 403 iterations. (c) Initial (solid) and modified (dashed) area functions. (d) Frequency response functions of the initial (solid) and final (dashed) area functions; the vertical lines represent the target formant frequencies.

$$\mathrm{F}_{n_0} = 800, 1136, 2770, 3448, 4240 \text{ Hz}, \quad n = 1,2,3,4,5 \tag{16}$$

but the target values were set to be,

$$\mathcal{F}_n = 800, 1136, \mathbf{2800}, \mathbf{3250}, \mathbf{3700} \text{ Hz}, \quad n = 1,2,3,4,5, \tag{17}$$

which specifies that the distance between the upper formants (F3, F4, and F5) will be decreased.

The perturbation pattern at the first iteration is shown in Fig. 3(a), and the converging z coefficients for F1, F2, and F3 are plotted in Fig. 3(b). In this case, 403 iterations were required so that $\delta < 0.1$ Hz. Shown in Fig. 3(c) are the initial and modified area functions. Relative to the initial configuration, the modifications consist of slight constrictions and expansions along the entire vocal tract length. The most apparent changes are in the oral portion of the area function, but the small changes between the glottis and about 8 cm may have also contributed significantly to the acoustic changes. The frequency response functions in Fig. 3(d) indicate that the target formant frequencies are achieved with the modified area function. F1 and F2 are the same as in the initial configuration, but F3, F4, and F5 were moved toward each other, and their combined effect produces a somewhat enhanced amplitude in the 2800–3800 Hz range. It can be noted parenthetically that the cluster of F3, F4, and F5 was produced without any major changes to the length and area of the epilaryngeal portion of the area function. This contrasts with the typical articulatory interpretation of the singing formant (Sundberg, 1974).

V. DISCUSSION

A technique has been proposed that allows area functions to be modified so that their formant frequencies match a set of targets. In both of the examples, the technique was used to successfully modify an original area function, reducing the difference between a set of original and target formant frequencies to nearly zero. An inherent limitation of the technique, however, is that it is not possible to know if the resulting modifications to an initial area function are those that would actually be produced by a human speaker. Hence, the modifications can only be considered hypothetical. Nonetheless, the ability to systematically perturb an area function such that a desired formant pattern is produced may be a useful tool to help understand how relatively subtle changes in cross-sectional area can produce significant acoustic changes.

ACKNOWLEDGMENTS

This study was supported by Grant No. R01 DC04789 from the NIDCD.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (**1978**). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer sorting-sorting technique," J. Acoust. Soc. Am. **63**, 1535–1555.

Carré, R. (**2004**). "From an acoustic tube to speech production," Speech Commun. **42**, 227–240.

Dang, J., and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. **101**(1), 456–465.

Espy-Wilson, C. Y. (**1992**). "Acoustic measures for linguistic features distinguishing the semivowels/w j r l/ in American English," J. Acoust. Soc. Am. **92**, 736–757.

Fant, G., and Pauli, S. (**1975**). "Spatial characteristics of vocal tract resonance modes," in Proceedings of the Speech Comm. Sem. 74., Stockholm, Sweden, August 1–3, pp. 121–132.

Makarov, I. S., and Sorokin, V. N. (**2004**). "Resonances of a branched vocal tract with compliant walls," Acoust. Phys. **50**(3), 323–330.

Mermelstein, P. (**1967**). "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Am. **41**(5), 1283–1294.

Milenkovic, P. (**1984**). "Vocal tract area functions from two-point acoustic measurements with formant frequency constraints," IEEE Trans. Acoust., Speech, Signal Process., **ASSP-32**(6), 1122–1135.

Schroeder, M. R. (**1967**). "Determination of the geometry of the human vocal tract by acoustic measurements," J. Acoust. Soc. Am. **41**(4), 1002–1010.

Sondhi, M. M., and Resnick, J. R. (**1983**). "The inverse problem for the vocal tract: Numerical methods, acoustical experiments, and speech synthesis," J. Acoust. Soc. Am. **73**(3), 985–1002.

Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**(7), 955–967.

Sorokin, V. N. (**1992**). "Determination of vocal tract shape for vowels," Speech Commun. **11**, 71–85.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**(1), 537–554.

Story, B. H., Laukkanen, A.-M., and Titze, I. R. (**2000**). "Acoustic impedance of an artificially lengthened and constricted vocal tract," J. Voice **14**(4), 455–469.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**2001**). "The relationship of vocal tract shape to three voice qualities," J. Acoust. Soc. Am. **109**, 1651–1667.

Sundberg, J. (**1974**). "Articulatory interpretation of the singing formant," J. Acoust. Soc. Am. **55**, 838–844.

Wakita, H. (**1973**). "Direct estimation of vocal tract shape by inverse filtering of acoustic speech waveforms," IEEE Trans. Audio Electroacoust. **AU-21**(5), 417–427.

Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period (L)[a)]

Diana Deutsch[b)]
Department of Psychology, University of California, San Diego, La Jolla, California 92093

Trevor Henthorn
Department of Music, University of California, San Diego, La Jolla, California 92093

Elizabeth Marvin
Department of Music Theory, Eastman School of Music, Rochester, New York 14604

HongShuai Xu
College of Music, Capital Normal University, Beijing, 10037, China

(Received 12 September 2005; revised 20 November 2005; accepted 21 November 2005)

Absolute pitch is extremely rare in the U.S. and Europe; this rarity has so far been unexplained. This paper reports a substantial difference in the prevalence of absolute pitch in two normal populations, in a large-scale study employing an on-site test, without self-selection from within the target populations. Music conservatory students in the U.S. and China were tested. The Chinese subjects spoke the tone language Mandarin, in which pitch is involved in conveying the meaning of words. The American subjects were nontone language speakers. The earlier the age of onset of musical training, the greater the prevalence of absolute pitch; however, its prevalence was far greater among the Chinese than the U.S. students for each level of age of onset of musical training. The findings suggest that the potential for acquiring absolute pitch may be universal, and may be realized by enabling infants to associate pitches with verbal labels during the critical period for acquisition of features of their native language. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151799]

PACS number(s): 43.75.Cd [NHF] Pages: 719–722

I. INTRODUCTION

Absolute pitch—the ability to name or produce a note of particular pitch in the absence of a reference note—is extremely rare in the U.S. and Europe, with an estimated prevalence in the general population of less than one in 10 000 (Profita and Bidder, 1988). Because of its rarity, and because many famous musicians have been known to possess it, absolute pitch is often considered to be a sign of exceptional musical ability. However, it is not necessarily accompanied by superior processing on other musical tasks (Lockhead and Byrd, 1981; Rakowski and Morawska-Büngeler, 1987; Takeuchi and Hulse, 1993; Burns and Campbell, 1994, Miyazaki and Rakowski, 2002). Furthermore, most people who do not have absolute pitch can nevertheless display surprisingly accurate long-term memories for pitch when verbal labeling is not required (Deutsch *et al.*, 1987; Levitin, 1994). The genesis of absolute pitch therefore presents a challenge for auditory theory.

Deutsch *et al.* (2004) found that speakers of two tone languages—Vietnamese and Mandarin—exhibited a remarkably precise and stable form of absolute pitch in reciting lists of words. Given these findings, it was conjectured that absolute pitch evolved as a feature of speech, analogous to other features such as vowel quality. [In tone languages, words take on entirely different meanings depending on the *tones* in which they are enunciated. Tones are defined both by their pitch heights ("registers") and by their pitch contours. For example, in Mandarin the word "ma" means "mother" when spoken in the first tone, "hemp" when spoken in the second tone, "horse" when spoken in the third tone, and a reproach when spoken in the fourth tone.] It was further conjectured that tone language speakers acquire this feature in infancy, during the critical period in which infants acquire other features of their native language (Kuhl *et al.*, 1992; Doupe and Kuhl, 1999) and so can later acquire absolute pitch for music in the same way as they acquire the features of a second tone language. This in turn led to the hypothesis that there should be a much higher prevalence of absolute pitch for music among tone language speakers than among speakers of nontone languages such as English. The acquisition of absolute pitch by rare individuals who are nontone language speakers was hypothesized to be associated with a critical period of unusually long duration, so that it encompasses the age at which the child can begin taking music lessons.

The present study examined the hypothesis of a speech-related critical period for acquiring absolute pitch, by comparing its prevalence in two large groups of music students. An on-site test was administered to the two groups under

[a)]Portions of this work were presented in "Tone language and absolute pitch: Prevalence among American and Chinese conservatory students," 148th Meeting of the *Acoustical Society of America*, San Diego, CA, November, 2004.

[b)]Electronic mail: ddeutsch@ucsd.edu

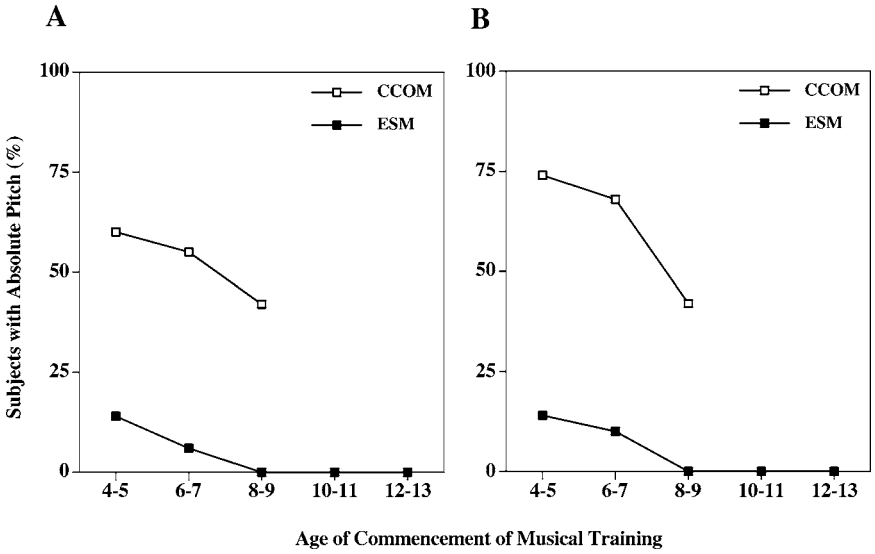


FIG. 1. Percentages of subjects who obtained a score of at least 85% correct on the test of absolute pitch, as a function of age of onset of musical training. (**A**) No semitone errors allowed; (**B**) semitone errors allowed. Unfilled boxes show the results from students at the Central Conservatory of Music (CCOM) in Beijing, China; these were all tone language speakers. Filled boxes show the results from students at Eastman School of Music (ESM), Rochester, NY, who were nontone language speakers.

highly similar conditions. The first group consisted of first-year students who were enrolled in a required course at the Central Conservatory of Music (CCOM) in Beijing; all these subjects spoke the tone language Mandarin. The second group consisted of first-year students who were enrolled in a required course at Eastman School of Music (ESM) and who were nontone language speakers, with both parents being nontone language speakers (see below).

II. METHOD

The CCOM group consisted of 88 subjects who all spoke Mandarin. These were 28 male and 60 female, with a mean age of 20 (range 17–34) years. The ESM group consisted of 115 subjects, 54 male and 61 female, with a mean age of 19 (range 17–23) years; these were all nontone language speakers, with both parents being nontone language speakers. (The data from an additional 18 students at ESM were excluded from analysis on the grounds that they, or at least one of their parents, spoke an Asian language. This was the case for all those students who described their ethnic background as Asian.) In both groups, all students who were invited to take the test agreed to do so; there was therefore no self-selection of subjects within either group. The two groups were subdivided into subgroups by age of onset of musical training. In order to make meaningful comparisons, only those subgroups that contained at least nine subjects were considered. These were CCOM: ages 4–5, $n=43$; ages 6–7, $n=22$; ages 8–9, $n=12$; ESM: ages 4–5, $n=21$; ages 6–7, $n=31$; ages 8–9, $n=24$; ages 10–11, $n=20$; ages 12–13; $n=9$.

The subjects were presented with a set of 36 notes that spanned the three-octave range from C_3 (131 Hz) to B_5 (988 Hz), and they were asked to indicate the name of each note in writing. In order to minimize the use of relative pitch as a cue, all intervals between successively presented notes were larger than an octave. The notes were piano tones generated on a Kurzweil K2000 synthesizer which was tuned to the standard A4 of 440 Hz, and were 500 ms in duration. They were presented in three blocks of 12, with 4.25 s intervals between onsets of notes within a block, and 39-s rest periods between blocks. The test blocks were preceded by a practice block of four notes. No feedback was provided, either during the practice block or during the test blocks. The notes were played to subjects via a CD or DVD player, amplifier, and two loudspeakers. The subjects were also asked to fill out a questionnaire concerning their music education, where they and their parents had lived, the languages they and their parents spoke, and (for the ESM group) what they considered to be their primary ethnic background.

III. RESULTS

Two separate criteria for absolute pitch were employed: The first consisted of obtaining a score of at least 85% correct on the test, and the second also consisted of obtaining a score of at least 85% correct, allowing for semitone errors. The data were analyzed separately according to both criteria.

The main results of the experiment are shown in Fig. 1, which displays two major findings. First, both the CCOM and the ESM groups showed orderly effects of age of onset of musical training: The earlier the age of onset, the higher the probability of meeting the criteria for absolute pitch. Second, the prevalence of absolute pitch was far higher for the CCOM than for the ESM group. Statistical analyses on the numbers of subjects meeting the criteria for absolute pitch showed overall effects of age of onset of musical training (with no semitone errors allowed: $\chi^2=11.44$, df=2, $p<0.01$; with semitone errors allowed, $\chi^2=16.85$, df=2, $p<0.001$). Comparisons between the CCOM and ESM groups were therefore made separately for each level of age of onset of musical training (Fisher Exact Probability Tests were used for all remaining comparisons). All comparisons were highly significant: With no semitone errors allowed, for those who began musical training at ages 4–5, $p<0.001$; at ages 6–7, $p<0.001$; and at ages 8–9, $p<0.005$. With semitone errors allowed, for those who began musical training at ages 4–5, $p<0.001$; at ages 6–7, $p<0.001$; and at ages 8–9, $p<0.005$. Comparisons were also made between the male and female subjects within each subgroup (six comparisons with no semitone errors allowed; six comparisons with semitone

errors allowed). No effects of gender were found ($p > 0.05$ for all comparisons, and there was no overall trend based on gender in either direction).[1]

IV. DISCUSSION

The present findings support the hypothesis that, if given the opportunity, infants can acquire absolute pitch as a feature of speech, which can later generalize to musical tones. Indeed, the plots shown here for the acquisition of absolute pitch in nontone language and tone language speakers reflect a very similar picture, in terms of time frame, to the critical periods inferred by other researchers for the acquisition of first and second languages, respectively (Johnson and Newport, 1989; Newport, 1990; Bates, 1992; Doupe and Kuhl, 1999). It should be noted that although other critical periods have been documented, for example for the development of ocular dominance columns in the visual cortex of cats (Hubel and Wiesel, 1970), and for auditory localization in barn owls (Knudsen, 1988), no other critical periods have been identified that show a similar correspondence with speech and language in terms of time frame.

Further supporting evidence for a speech-related critical period hypothesis comes from the finding that 8-month-old infants can perform perceptual learning tasks that require referring to the absolute pitches of tones (Saffran and Griepentrog, 2001), indicating that preverbal infants have the capacity to acquire absolute pitch. Other evidence linking absolute pitch to speech comes from structural MRI findings that musicians with absolute pitch have an exaggeration of the normal leftward asymmetry of the planum temporale—an area that is considered to be critically involved in speech processing (Schlaug *et al.*, 1995). As a further indication, the probability of acquiring absolute pitch has been reported to be inversely related to the age of onset of musical training (Baharloo *et al.*, 1998; Profita and Bidder, 1988), suggesting a critical period for its acquisition.

The present findings should also be considered in relation to the results of surveys obtained by Gregersen *et al.* (1999, 2000) involving students taking music classes within the U.S. The responses indicated a higher prevalence of absolute pitch among those who described their ethnic background as "Asian" than among "nonAsians." Although language was not considered in these studies, the majority of respondents in the Asian category had been born in Asia, and their primary language would therefore have been a tone language or a pitch accent language.[2] Furthermore, of those born in the U.S., a large number would have had an Asian language as their first language. Because the data obtained by Gregersen *et al.* (1999, 2000) were obtained from surveys—and so involved self-selection and were based on self-report—and because the reported prevalence of absolute pitch was not documented as a function of age of onset of musical training for the different Asian linguistic groups taken separately, their relationships to the present findings are unclear. Nevertheless, it may alternatively be proposed that the differences between the two groups obtained in the present study were due to genetic factors (see also Baharloo *et al.*, 1998). Indeed, both critical period and genetic factors might be involved, since there may be strong selective pressure among tone language speakers to possess absolute pitch (Deutsch *et al.*, 2004). In any event, this paper reports the first large-scale study which demonstrates a difference in the prevalence of absolute pitch in two normal populations, as determined by administration of an on-site test, without self-selection from within the target populations, and controlling for gender and for age of onset of musical training.

ACKNOWLEDGMENTS

The authors wish to thank Zhou Haihong, Zhang Zheng-Sheng, Steve Laitz, W. Marvin, C. Harris, K. Hamaoui, and D. Bosnjakovic for help in different phases of the study, and E. Newport for valuable discussions.

[1]Some semitone errors may have resulted, in part, from experience with different scales and tuning systems, particularly in childhood. This would be expected to apply more, in general, to those subjects who were brought up on Asian musical scales.

[2]Many Asian languages, for example, Mandarin, Cantonese, Vietnamese, and Thai are tonal, and certain others, for example Japanese and certain dialects of Korean, are pitch accent; i.e, the pitch structure of certain words is used to differentiate meaning (Pickett, 1999).

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., and Freimer, N. B. (**1998**). "Absolute pitch: An approach for identification of genetic and nongenetic components," Am. J. Hum. Genet. **62**, 224–231.

Bates, E. (**1992**). "Language development," Curr. Opin. Neurobiol. **2**, 180–185.

Burns, E. M., and Campbell, S. L. (**1994**). "Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception?" J. Acoust. Soc. Am. **96**, 2704–2719.

Deutsch, D., Henthorn, T., and Dolson, M. (**2004**). "Absolute pitch, speech, and tone language: Some experiments and a proposed framework," Music Percept. **21**, 339–356.

Deutsch, D., Kuyper, W. L., and Fisher, Y. (**1987**). "The tritone paradox: Its presence and form of distribution in a general population," Music Percept. **5**, 79–92.

Doupe, A. J., and Kuhl, P. K. (**1999**). "Birdsong and human speech: Common themes and mechanisms," Annu. Rev. Neurosci. **22**, 567–631.

Gregersen, P. K., Kowalksy, E., Kohn, N., and Marvin, E. W. (**1999**). "Absolute pitch: Prevalence, ethnic variation, and estimation of the genetic component," Am. J. Hum. Genet. **65**, 911–913.

Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. (**2000**). "Early childhood music education and predisposition to absolute pitch," Am. J. Med. Genet. **98**, 280–282.

Hubel, D. H., and Wiesel, T. N. (**1970**). "The period of susceptibility to the physiological effects of unilateral eye closure in kittens," J. Physiol. (London) **206**, 419–436.

Johnson, J. S., and Newport, E. L. (**1989**). "Critical periods in second language learning: The influence of maturational state on the acquisition of English as a second language," Cogn. Psychol. **21**, 60–99.

Knudsen, E. I. (**1988**). "Sensitive and critical periods in the development of sound localization," in *From Message to Mind: Directions in Developmental Neurobiology*, edited by S. S. Easter, K. F. Barald, and B. M. Carlson (Sinauer Associates, Sunderland, MA).

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.

Levitin, D. J. (**1994**). "Absolute memory for musical pitch: Evidence for the production of learned melodies," Percept. Psychophys. **56**, 414–423.

Lockhead, G. R., and Byrd, R. (**1981**). "Practically perfect pitch," J. Acoust. Soc. Am. **70**, 387–389.

Miyazaki, K., and Rakowski, A. (**2002**). "Recognition of notated melodies by possessors and nonpossessors of absolute pitch," Percept. Psychophys. **64**, 1337–1345.

Newport, E. L. (**1990**). "Maturational constraints on language learning," Cogn. Sci. **14**, 11–28.

Pickett, J. M. (**1999**). *The Acoustics of Speech Communication: Fundamen-*

tals, Speech Perception Theory, and Technology (Allyn and Bacon, Needham Heights).
Profita, J., and Bidder, T. G. (**1988**). "Perfect pitch," Am. J. Med. Genet. **29**, 763–771.
Rakowski, A., and Morawska-Büngeler, M. (**1987**). "In search of the criteria for absolute pitch," Arch. Acoust. **12**, 75–87.
Saffran, J. R., and Griepentrog, G. J. (**2001**). "Absolute pitch in infant auditory learning: Evidence for developmental reorganization," Dev. Psychol. **37**, 74–85.
Schlaug, G., Jaencke, L., Huang, Y., and Steinmetz, H. (**1995**). "*In vivo* evidence of structural brain asymmetry in musicians," Science **267**, 699–701.
Takeuchi, A. H., and Hulse, S. H. (**1993**). "Absolute pitch," Psychol. Bull. **113**, 345–361.

Description and measurement of the acoustic characteristics of two-tailpipe mufflers

A. J. Torregrosa,[a] A. Broatch, and T. Fernández
CMT-Motores Térmicos, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022-Valencia, Spain

F. D. Denia
Departamento de Ingeniería Mecánica y de Materiales, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022-Valencia, Spain

(Received 17 March 2005; revised 31 October 2005; accepted 21 November 2005)

Two-tailpipe mufflers have become a usual solution for flow noise abatement in turbocharged Diesel engines, since the mean flow velocity at the exhaust outlet may be reduced without increasing the tailpipe diameter, which would give rise to an increase of noise associated with engine orders. In this communication, the acoustic characteristics of mufflers with two tailpipes are studied. First, a global representation of the muffler acoustics in terms of three reflection coefficients and six transmission coefficients is presented, and a procedure to obtain, from such representation, a conventional two-port transfer matrix by setting the reflection coefficient at one of the outlet pipes is provided. A procedure based on the extension of the impulse method as applied to single inlet-single outlet mufflers was devised for the experimental determination of the transmission and reflection coefficients, and experiments on a nonsymmetric oval expansion chamber allowed one to check the suitability of the representation. The results indicate that nonsymmetries in the location of the tailpipes may be relevant for the design of realistic mufflers, as they may affect the source located at each tailpipe outlet and thus the resulting exhaust noise radiated. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159228]

PACS number(s): 43.20.Ye, 43.50.Gf [SFW] Pages: 723–728

I. INTRODUCTION

The increasing importance of flow noise, mainly in the case of turbocharged engines in which very high flow rates are required in order to achieve current power and torque design values, has made it necessary to consider means to reduce flow noise while keeping under control the pulsating component of exhaust noise.[1] The obvious solution to flow noise excess, which would be simply to increase the area of the tailpipe with the corresponding reduction in the velocity of the discharged jet,[2] is not a suitable one, since an increase in the discharge section induces an increment in the radiated pulsating noise.[3] A convenient alternative is to introduce two tailpipes in parallel, which has become common practice. This produces a decrease in the mean velocity of the exhaust jet, while keeping the same diameter, so that pulsating noise is not necessarily increased. The only point to consider is that, since a single source has been replaced by two sources, both the flow field in the discharge region and the radiation field at a certain distance may be modified. These eventual modifications must, in principle, depend on the relative position of the two tailpipes, but also on any eventual differences between the noise radiated by each of the tailpipes. Therefore, proper design of these mufflers should account for such differences, and thus a suitable representation which allows one to account for these issues is required.

Additionally, two-tailpipe mufflers present an inconvenience from the point of view of their inclusion into usual linear calculation methods, since these use a two-port representation for each muffler and for the exhaust line as a whole. Therefore, the problem is twofold: (i) to find a suitable representation which may give full information about the acoustic behavior of a two-tailpipe muffler, and (ii) to obtain, from this full characterization, a two-port representation that may allow for the inclusion of such information into usual calculation tools and for the analysis and comparison of the results.

These two issues are addressed in the present paper. First, in Sec. II, the theory of the representation proposed will be developed and the experimental setup used will be described. Then, experimental results for a simple but representative case will be shown and discussed in Sec. III. Finally, in Sec. IV, a summary of the work presented together with some conclusions are given.

II. BACKGROUND

While several formalisms may be used for the representation of the acoustic response of a muffler,[4] the most intuitive one for the problem under study is that based on the consideration of wave components instead of pressure and velocity fluctuations. Considering a conventional two-port muffler one may assume, as shown in Fig. 1(a), that both upstream and downstream of the muffler there are two pressure components: one moving toward the muffler and the other moving away from the muffler. Denote by A the components moving toward the muffler, and by B the components moving away. Denote also by 1 the upstream section

[a]Electronic mail: atorreg@mot.upv.es

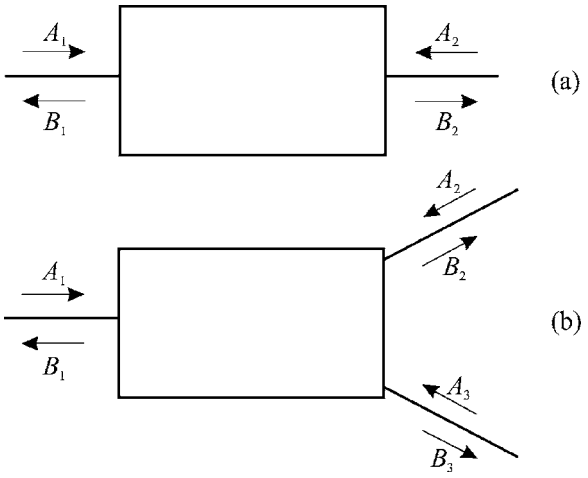


FIG. 1. Pressure components in: (a) two-port system, (b) three-port system.

and by 2 the downstream section. Then, the following relation can be established between the excitations A_i and the muffler responses B_i:

$$B_1 = R_1 A_1 + T_{21} A_2,$$
$$B_2 = R_2 A_2 + T_{12} A_1 \tag{1}$$

or, in compact matrix form,

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} R_1 & T_{21} \\ T_{12} & R_2 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}. \tag{2}$$

Here, R_i represents the reflection coefficient as seen from port i, and T_{ij} the transmission coefficient from port i to port j. This notation allows for the direct extension to a two-tailpipe muffler, just by not considering it as a single input-double output device (as it works in practice), but as a general three-port element. In this way, for the element represented in Fig. 1(b), one has three excitations and three responses, and writing the relations between them directly in matrix form, one has[5]

$$\begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} R_1 & T_{21} & T_{31} \\ T_{12} & R_2 & T_{32} \\ T_{13} & T_{23} & R_3 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}. \tag{3}$$

In this way, one has a reflection coefficient for each of the pipes arriving at the muffler, and transmission coefficients for all the possible transmission paths, indicated by the corresponding subscripts. The knowledge of these nine complex coefficients gives the full representation of the acoustic behavior of the muffler. In order to determine their values, the three following measurements are required.

(1) Excitation in duct 1, with anechoic terminations in ducts 2 and 3, so that $A_1 \neq 0$ and $A_2 = A_3 = 0$, and thus

$$R_1 = B_1/A_1, \quad T_{12} = B_2/A_1, \quad T_{13} = B_3/A_1. \tag{4}$$

(2) Excitation in duct 2, with anechoic terminations in ducts 1 and 3, so that $A_2 \neq 0$ and $A_1 = A_3 = 0$; then

$$R_2 = B_2/A_2, \quad T_{21} = B_1/A_2, \quad T_{23} = B_3/A_2. \tag{5}$$

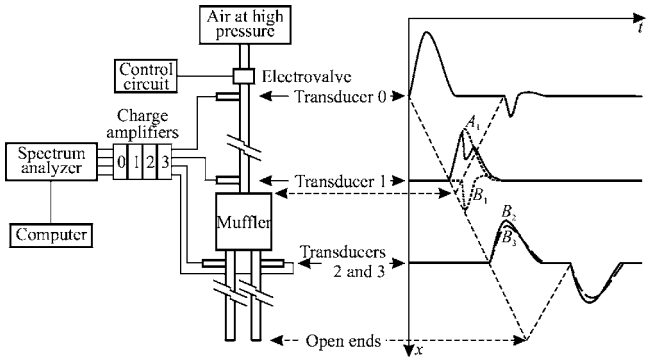


FIG. 2. Scheme of the experimental setup and illustration of relevant pressure components.

(3) Excitation in duct 3, with anechoic terminations in ducts 1 and 2; now, one has $A_3 \neq 0$ and $A_1 = A_2 = 0$, and therefore

$$R_3 = B_3/A_3, \quad T_{31} = B_1/A_3, \quad T_{32} = B_2/A_3. \tag{6}$$

Different experimental techniques are available (see, for instance, Ref. 6) in order to perform the above-indicated tests. In the present work, a modified version of the impulse method[7] was used, since pressure components, on which all the previous developments are based upon, are directly obtained in the time domain with a simple procedure, and the extension of its use to the case of three-port elements is straightforward. In Fig. 2 both the modified experimental setup used and the relevant pressure waves recorded are illustrated.

The incident pulse is generated by means of a high-speed electrovalve that controls the discharge from a high pressure reservoir. A proper choice of the opening time ensures that the spectrum associated with the incident pulse is essentially flat. The length of the ducts placed between the valve and transducer 0, transducer 0 and the muffler, and the muffler and the open ends is chosen so that no windowing is necessary in order to isolate the incident, the reflected, and the transmitted pulses, as indicated in the figure. Transducer 0 was located 15 m away from both the valve and the muffler, and transducers 2 and 3 were placed 0.1 m downstream of the muffler, and 15 m away from their corresponding open end. However, information on reflection properties as recorded by transducer 0 is highly distorted, due precisely to the huge duct length required to avoid overlapping. This information is only available at points located close to the muffler, such as that indicated for transducer 1 in Fig. 2, where it is also indicated that this transducer records the addition of the incident and the reflected pulses. In order to surpass this difficulty, the solution adopted in Ref. 7 was to estimate the pulse incident on the muffler at section 1 (whose Fourier transform will give the complex amplitude of the A_1 component) from an additional test performed without any muffler, and using the pressure recorded by transducer 0 only to check the comparability of the excitations used in both types of tests (with and without muffler). Once the incident pulse is so obtained, subtraction from the pressure recorded by transducer 1 in the presence of a muffler gives a suitable estimate of the pulse reflected by the muffler, whose Fourier

transform gives the complex amplitude of the B_1 component. In this way, neither the incident nor the reflected pulse recorded by transducer 0 is used for the muffler characterization, so that any effects of wave distortion or of dissipation due to shear will not affect the results.

Transmitted pulses do not pose any problem, since they are obtained simply by rejecting the backward components coming from the open ends; again, by applying the Fourier transform to the "clean" signals one obtains the complex amplitude of the B_2 and B_3 components. Then, reflection coefficient R_1 and transmission coefficients T_{12} and T_{13} are obtained from Eq. (4). Two additional tests are thus required in order to determine the other two sets of reflection and transmission coefficients, in which the muffler must be excited from sections 2 and 3 and the coefficients are obtained from Eqs. (5) and (6), respectively.

Now, a complete characterization of the acoustic performance of the muffler is available, and thus the problem of transforming this 3×3 transmission matrix into a conventional 2×2 transmission matrix will be addressed. With this purpose, assume that duct 1 is the inlet to the muffler, and consider that duct 2 is the outlet. In this way, duct 3 may be described as a side branch, consisting of a duct of a certain length finished by an open termination. The behavior of this system is well known, so that one may write

$$A_3 = \tilde{R}_3 B_3, \tag{7}$$

where $\tilde{R}_3$ represents the reflection coefficient associated with the duct downstream of the measurement section together with the open end. Denoting by l_3 the length of this duct portion, one has[8]

$$\tilde{R}_3 = -R_0 \exp[-\mathrm{j}\beta 2(l_3 + \delta_0)]. \tag{8}$$

Here, $\mathrm{j}=(-1)^{1/2}$, $\beta \simeq k - \mathrm{j}\alpha$ is the propagation constant, with $k=2\pi f/a_0$, f being the frequency and a_0 the speed of sound, and α is the visco-thermal attenuation coefficient which, for plane wave propagation in a circular pipe of radius r, can be expressed as $\alpha=(1/ra_0)(\nu\pi f)^{1/2}[1+(\gamma-1)\mathrm{Pr}^{-1/2}]$, where ν is the kinematic viscosity, $\gamma=c_p/c_v$ the ratio of the specific heats, and Pr is the Prandtl number. Finally, the modulus R_0 of the reflection coefficient for an unflanged open end and the end correction δ_0 representing its phase may be closely approximated by[8]

$$R_0 = 1 + 0.013\,36kr - 0.590\,79(kr)^2 + 0.335\,76(kr)^3 - 0.6432(kr)^4, \quad 0 < kr < 1.5 \tag{9}$$

and

$$\delta_0/r = \begin{cases} 0.6133 - 0.1168(kr)^2, & kr < 0.5 \\ 0.6393 - 0.1104kr, & 0.5 < kr < 2, \end{cases} \tag{10}$$

respectively.

Now, from Eq. (3), one has

$$B_3 = T_{13}A_1 + T_{23}A_2 + R_3A_3 \tag{11}$$

and making use of Eq. (7) one may write

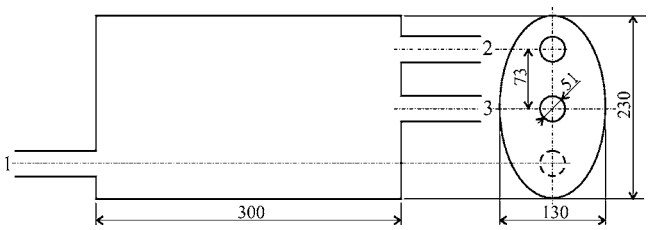


FIG. 3. Expansion chamber tested (dimensions in mm).

$$A_3 = \frac{T_{13}A_1 + T_{23}A_2}{\vartheta_3} \quad \text{with } \vartheta_3 = \frac{1}{\tilde{R}_3} - R_3 \tag{12}$$

with $\tilde{R}_3$ given by Eq. (8), so that A_3 is expressed in terms of known quantities. Therefore, the three equations representing the acoustic behavior of the muffler reduce to the two following ones:

$$\begin{aligned} B_1 &= [R_1 + (T_{13}T_{31}/\vartheta_3)]A_1 + [T_{21} + (T_{23}T_{31}/\vartheta_3)]A_2, \\ B_2 &= [T_{12} + (T_{13}T_{32}/\vartheta_3)]A_1 + [R_2 + (T_{23}T_{32}/\vartheta_3)]A_2. \end{aligned} \tag{13}$$

Finally, comparison with Eq. (1) allows the identification of the transmission and reflection coefficients of the equivalent two-port muffler, as

$$\begin{aligned} \hat{R}_1 &= R_1 + (T_{13}T_{31}/\vartheta_3), \quad \hat{T}_{12} = T_{12} + (T_{13}T_{32}/\vartheta_3), \\ \hat{R}_2 &= R_2 + (T_{23}T_{32}/\vartheta_3), \quad \hat{T}_{21} = T_{21} + (T_{23}T_{31}/\vartheta_3). \end{aligned} \tag{14}$$

Of course, if necessary this information may be rendered into the form of a usual transfer matrix,[4] as required by most acoustic calculation tools.

III. EXPERIMENTS

In order to check the performance of the proposed representation, a simple expansion chamber with the geometry shown in Fig. 3 was considered. The acoustics of expansion chambers with two outlets is well known: it has been studied in Ref. 9 in the case of circular cross section, and in Ref. 10 in the case of elliptical cross section. Apart from the essential simplicity of the geometry, which gives a good test case for the consistency of the method, this case was chosen in order to account for differences between the two transmission paths. The fact that both the inlet and one of the two outlets of this chamber are not centered should give information about the eventual importance of tailpipe location, since transversal modes are expected to appear and they will affect the wave transmitted to the two outlets in a different way. To obtain a first assessment of these issues, in Figs. 4–6 and 8 the amplitude of the measured transmission and reflection coefficients is plotted as a function of the normalized frequency f/a_0. In order to facilitate the analysis, the transmission coefficients are plotted by pairs.

First, in Fig. 4, T_{12} and T_{13} are represented. It can be observed that in T_{12} a sharp rise in transmission can be seen at around $f/a_0=2.5$, which corresponds to the typical collapse in attenuation associated with transversal modes, while this does not appear in the case of T_{13}. Obviously, this is due to the fact that the positions of pipes 2 and 3 relative to pipe

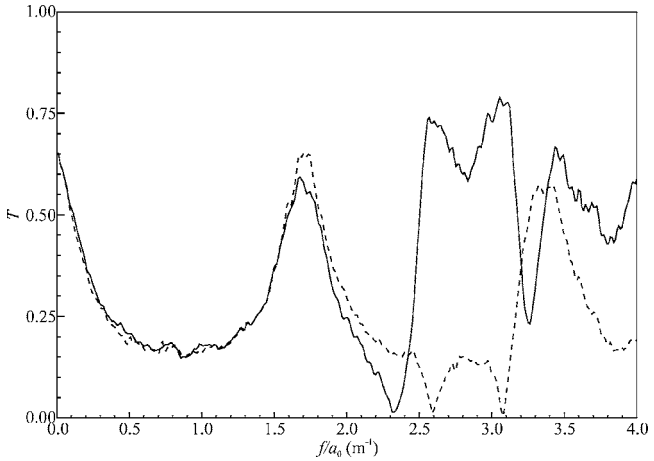


FIG. 4. Measured results: comparison of transmission coefficients T_{12} (solid line) and T_{13} (dashed line).

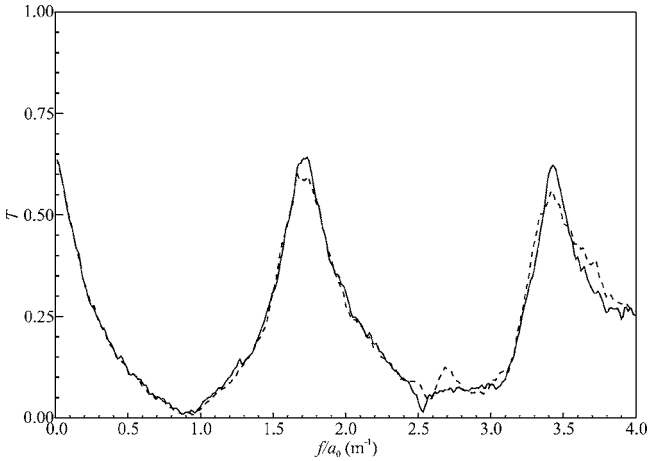


FIG. 6. Measured results: comparison of transmission coefficients T_{23} (solid line) and T_{32} (dashed line).

1 are different, and hence transmission is affected. The same feature is observed in Fig. 5, where the corresponding inverse coefficients are plotted. Again, when pipe 2 is involved a transversal mode effect appears, as evidenced by the sharp rise observed in T_{21}. On the contrary, as shown in Fig. 6, the transmission coefficients between pipes 2 and 3 do not present any remarkable differences. In fact, comparison of Figs. 5 and 6 indicates that the characteristic frequencies in the attenuation patterns of T_{31} and T_{32} are roughly the same, being determined by the fact that pipe 3 is centered and pipes 1 and 2 are located at the same distance from the muffler axis. The only remarkable difference between T_{31} and T_{32} can be seen around $f/a_0=0.9$, where T_{31} has an attenuation maximum with the "dome" shape typical of expansion chambers, whereas in T_{32} a clear attenuation spike is observed which reasonably resembles the behavior of an equivalent reversal chamber.[11]

These results are fully consistent with the transversal mode shapes for an elliptical chamber.[11] The first transversal mode to be excited is the even mode (1,0), symmetric with respect to the major axis and antisymmetric with respect to the minor axis, with the nodal line located along the minor axis. For eccentricities $\varepsilon > 0.05$ the cut-off frequency f_c for this mode is given by[11]

$$(f_c\delta/a_0)_{1,0} = 7.0771 \times 10^{-7} + 0.293\varepsilon + 2.1415 \times 10^{-4}\varepsilon^2 + 0.005\varepsilon^3 + 0.0021\varepsilon^4 - 0.0013\varepsilon^5 + 0.0012\varepsilon^6, \quad (15)$$

where δ is the semi-interfocal distance. In the present case, from the dimensions shown in Fig. 3 one has $\varepsilon=0.8249$ and $\delta=0.0948$ m, and substitution into Eq. (12) gives $f_c/a_0=2.55$, which is in perfect agreement with the behavior observed. The shape of the (1,0) mode is shown in Fig. 7, where it is apparent that pipes 2 and 3 are affected in quite a different way, thus justifying the differences observed in the associated transmission paths.

With respect to the reflection coefficients depicted in Fig. 8, it can be observed that, for frequencies below the aforementioned cut-off frequency, only local differences are found, the shape of the curves being approximately that found for a concentric expansion chamber.[7] Above that frequency, it is observed that, while R_1 and R_2 still have approximately the same shape, with differences in the levels and a certain shift in the characteristic frequencies, the be-

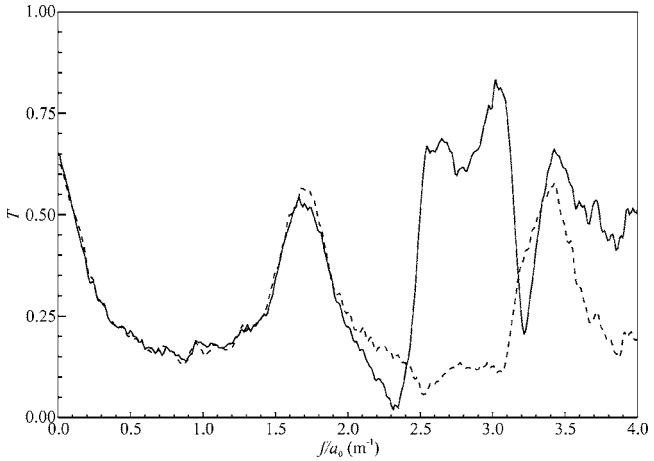


FIG. 5. Measured results: comparison of transmission coefficients T_{21} (solid line) and T_{31} (dashed line).

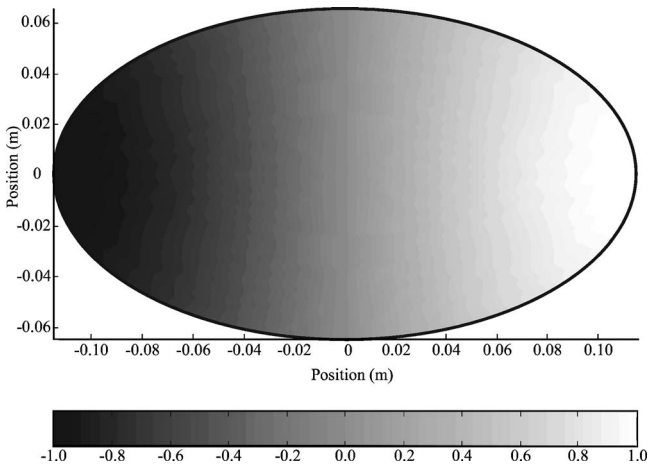


FIG. 7. Transversal pressure distribution for the (1,0) mode.

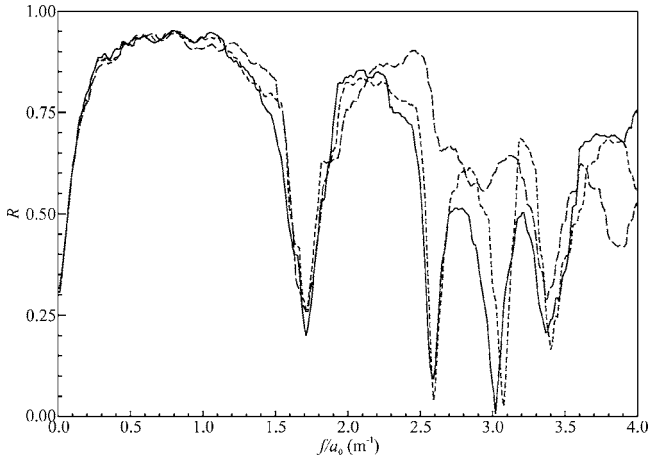


FIG. 8. Measured results: comparison of reflection coefficients R_1 (solid line), R_2 (short-dashed line), and R_3 (long-dashed line).

havior of R_3 is quite different for frequencies between $2.2a_0$ and $3.2a_0$; again, the explanation should lie in the relation between the distribution of the relevant transversal mode and the position of the pipe considered; in fact, from Fig. 3 it is intuitively clear that the muffler is quite different as seen from pipe 3, which is centered, than as seen from pipes 1 and 2.

Finally, the possibility to obtain a two-port representation, as indicated by Eq. (11), from the three-port representation available, is illustrated in Fig. 9. In this case, the same chamber but with one of the tailpipes open (that labeled as 2 in Fig. 3) was measured with the conventional arrangement of the experimental setup for single inlet and outlet. In all the cases, it can be seen that acceptable results are obtained: all the features found in the measured results are reproduced by those reconstructed from the three-port characterization. Only in the vicinity of abrupt changes of slope a certain amplitude difference and a small frequency shift can be observed, which are more apparent for f/a_0 between 1.5 and 1.7. Such differences should in principle be attributed to the representation of the open end given by Eqs. (9) and (10) and, in any case, are small, thus indicating the feasibility of

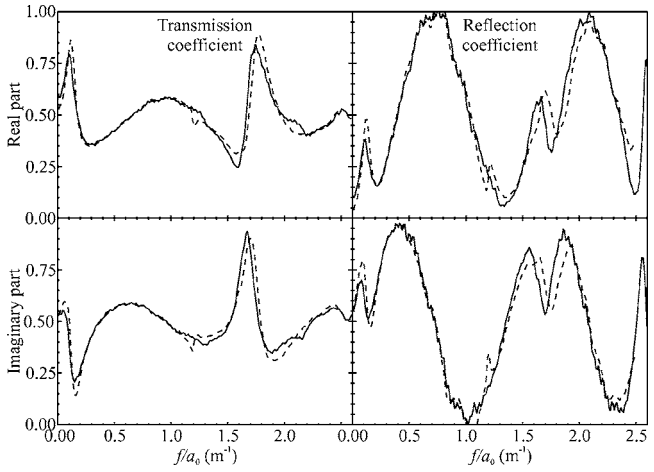


FIG. 9. Measurements with one open tailpipe: comparison of measured (solid line) and reconstructed from the three-port characterization (dashed line) transmission (left) and reflection (right) coefficients.

the procedure and hence the possibility to incorporate the information obtained into usual two-port calculations.

IV. CONCLUSION

A suitable description of the detailed acoustic behavior of mufflers with two tailpipes, in terms of three reflection coefficients and six transmission coefficients, has been presented. An experimental procedure for the determination of those coefficients has been implemented which is a natural extension of that associated with the impulse method as applied to single inlet-single outlet mufflers. This experimental procedure requires three tests, each yielding two transmission coefficients and one reflection coefficient, thus providing a global representation of the muffler acoustics. Then, the procedure to obtain, from such representation, a conventional two-port transfer matrix by setting the reflection characteristics at one of the branches has been presented.

Experimental check of the representation has been obtained through the study of a nonsymmetric oval expansion chamber with two outlets. The trends observed in the measured transmission and reflection coefficients have been found to be consistent with the expected behavior for such a chamber, and indicate the importance of eventual nonsymmetries in the location of the tailpipes in relation to the acoustic field distribution inside the chamber. This might be a key point in the design of more realistic mufflers, since such nonsymmetries will cause the equivalent noise sources at the tailpipe open ends to be different, thus affecting the resulting acoustic field radiated. More precisely, this indicates that when designing a muffler with two tailpipes it may not be advisable simply to duplicate the tailpipe geometry used in an existing design with a single tailpipe, without further consideration of the effects shown here.

An additional check of the consistency of the description has been obtained by performing a conventional two-port element characterization with the same chamber but keeping one of the tailpipes open, and comparing the results with those obtained from the three-port characterization and making suitable assumptions on the reflection coefficient of the open tailpipe. The reasonable agreement obtained suggests that the proposed procedure is able to provide relevant information on the behavior of two-tailpipe mufflers.

ACKNOWLEDGMENT

This work has been partially supported by Ministerio de Ciencia y Tecnología through Grant No. DPI2003-07153-C02-02.

[1]F. H. Kunz, "Semi-empirical model for flow noise prediction on intake and exhaust system," SAE Paper 1999-01-1654, Society of Automotive Engineers, Warrendale, PA, 1999.
[2]J. M. Desantes, A. J. Torregrosa, and A. Broatch, "Experiments on flow noise generation in simple exhaust geometries," Acta. Acust. Acust. **87**, 46–55 (2001).
[3]P. O. A. L. Davies, "Radiation of sound by a hot exhaust," J. Sound Vib. **76**, 591–594 (1981).
[4]P. O. A. L. Davies, "Transmission matrix representation of exhaust system acoustic characteristics," J. Sound Vib. **151**, 333–338 (1991).
[5]B. R. Mace, "Reciprocity, conservation of energy and some properties of reflection and transmission coefficients," J. Sound Vib. **155**, 375–381 (1991).

[6]M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987).
[7]F. Payri, J. M. Desantes, and A. Broatch, "Modified impulse method for the measurement of the frequency response of acoustic filters to weakly nonlinear transient excitations," J. Acoust. Soc. Am. **107**, 731–738 (2000).
[8]P. O. A. L. Davies, "Practical flow duct acoustics," J. Sound Vib. **124**, 91–115 (1988).
[9]A. Selamet and Z. L. Ji, "Acoustic attenuation performance of circular expansion chambers with single-inlet and double-outlet," J. Sound Vib. **229**, 3–19 (2000).
[10]F. D. Denia, L. Baeza, J. Albelda, and F. J. Fuenmayor, "Acoustic behavior of elliptical mufflers with single-inlet and double-outlet," in Proceedings of the Tenth International Congress on Sound and Vibration, Vol. **6**, Stockholm, Sweden, 2003, pp. 3287–3294.
[11]F. D. Denia, J. Albelda, F. J. Fuenmayor, and A. J. Torregrosa, "Acoustic behavior of elliptical chamber mufflers," J. Sound Vib. **241**, 401–421 (2001).

Elastoacoustic model with uncertain mechanical properties for ultrasonic wave velocity prediction: Application to cortical bone evaluation

Karina Macocco[a)]
Laboratoire de Mécanique Physique, CNRS UMR 7052 B2OA, Faculté des Sciences et Technologie, Université Paris XII-Val de Marne 61, Avenue du Général de Gaulle, 94010 Créteil Cédex, France

Quentin Grimal[b)]
Laboratoire d'Imagerie Paramétrique, CNRS UMR 7623 Université Pierre et Marie Curie 15, rue de l'école de médecine, 75006 Paris, France

Salah Naili[c)]
Laboratoire de Mécanique Physique, CNRS UMR 7052 B2OA, Faculté des Sciences et Technologie, Université Paris XII-Val de Marne 61, avenue du Général de Gaulle, 94010 Créteil Cédex, France

Christian Soize[d)]
Laboratoire de Mécanique, Université de Marne la Vallée, 5 boulevard Descartes, 77454 Marne la Vallée Cédex 2, France

(Received 11 August 2005; revised 9 November 2005; accepted 9 November 2005)

The axial transmission technique can measure the longitudinal wave velocity of an immersed solid. An elementary model of the technique is developed with a set of source and receivers placed in a semi-infinite fluid coupled at a plane interface with a semi-infinite solid. The acoustic fluid is homogeneous. The solid is homogeneous, isotropic, and linearly elastic. The work is focused on the prediction of the measured velocity (apparent velocity) when the solid is considered to have random material properties. The probability density functions of the random variables modeling each mechanical parameter of the solid are derived following the maximum entropy principle. Specific attention is paid to the modeling of Poisson's ratio so that the second-order moments of the velocities remain finite. The stochastic solver is based on a Monte Carlo numerical simulation and uses an exact semianalytic expression of the acoustic response derived with the Cagniard–de Hoop method. Results are presented for a solid with the material properties of cortical bone. The estimated mean values and confidence regions of the apparent velocity are presented for various dispersion levels of the random parameters. A sensibility analysis with respect to the source and receivers locations is presented. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146110]

PACS number(s): 43.20.Bi [RMW] Pages: 729–740

I. INTRODUCTION

In this paper we deal with predicting the reflection of a transient wave at a plane interface between a semi-infinite fluid and a semi-infinite solid in the ultrasonic range, when the source and receiver are placed in the fluid. This configuration is an elementary model of the "axial transmission" technique used to evaluate the mechanical properties of cortical bone [Bossy *et al.* (2004); Foldes *et al.* (1995); Lowet and Van der Perre (1996)]. This technique may also be used for non-destructive evaluation of classical engineering materials. The technique uses the lateral wave (also known as "head wave") that propagates in the solid at the velocity of longitudinal waves, close to the interface, and is refracted in the fluid at the critical angle [Brekhovskikh (1960)]. From the measurements of the times of flight associated with the lateral wave, the longitudinal wave velocity in the solid can be estimated; this estimation is referred to as the *apparent* velocity. In the context of bone evaluation, the solid and the fluid represent bone and soft tissues (skin, muscle), respectively. The purpose of such ultrasonic measurements of bones is to identify individuals with excessive bone fragility and fracture risk associated with various pathologies.

In principle, it would be possible to estimate mechanical parameters of the solid from the measured elastoacoustic response of the system that is built as the superposition of two semi-infinite media separated by a plane interface. This implies the solution of an intricate inverse problem. This work, which presents the forward modeling of the elastoacoustic response with several simplifying assumptions, is a first step toward the identification of the mechanical parameters with the axial transmission technique. This study should contribute to a better interpretation of measurements and help to develop new ultrasonic methods.

The novelty of this work is mainly to consider the solid homogeneous, but with random material properties. For in-

[a)]Electronic mail: macocco@univ-paris12.fr
[b)]Electronic mail: quentin.grimal@upmc.fr
[c)]Electronic mail: naili@univ-paris12.fr
[d)]Electronic mail: soize@univ-mlv.fr

stance, a homogeneous, linearly elastic, isotropic solid will be described by two elastic coefficients and the mass density, which are all assumed constant over the solid volume; however, the value of the elastic coefficients and of the mass density are not known in a deterministic way. In other words, we suppose that the solid under consideration is an element of an ensemble of homogeneous solids whose material properties obey probabilistic laws. The probabilistic framework is thought to be a good candidate to take into account intrinsic physiological variations of bone material properties. Indeed, bone can be viewed as a homogeneous material with random properties because its effective macroscopic properties depend on many factors (genetic, environmental, physiological, pathological), at various length scales.

The fluid parameters and the respective locations of the source and receivers are deterministic parameters in the problem.

The uncertainties of the solid material parameters are modeled by using the usual parametric approach, that is, each uncertain mechanical parameter is modeled by a random variable. Particular attention is devoted to the probabilistic model of the random variable associated with the Poisson ratio in order that the longitudinal and transverse wave velocities have a finite second-order moment. The probability models of the uncertain parameters are then constructed by using the maximum entropy principle [Shannon (1948); Jaynes (1957a,b); Kapur and Kesavan (1992)]. The maximum entropy principle is especially powerful in constructing a probabilistic model of bone because available experimental data sets are usually not sufficiently large to estimate a probability density functions by using mathematical statistics. The expressions derived in this paper for the probability density function are *a priori* valid for all bone types and can be adapted to any specific set of bone samples (specific bone type, specific species, etc.) by providing a mean value.

Taking advantage of the simple configuration considered, an exact solution of the elastoacoustics problem is used. This solution, corresponding to the reflection of a cylindrical wave on a semi-infinite solid, is obtained with the help of the Cagniard–de Hoop technique [Cagniard (1939); de Hoop (1960)]. The solution is available in a semianalytic form, namely closed-form analytic Green's function convolved with a function of the source history.

The final output of the method developed is the probability density function of the apparent velocity of longitudinal waves in the solid. It is calculated by using the probabilistic model associated with the elastoacoustic equations. The stochastic solver is based on the Monte Carlo simulation method. Realizations of the random variables are obtained with generators adapted to the problem. The computational time is reasonable thanks to the use of the Cagniard–de Hoop solution.

The paper is organized as follows. The elastoacoustic problem is formulated in Sec. II. In Sec. III, the procedure for calculating the apparent velocity is presented for the mean problem, that is, for the associated deterministic problem; the Cagniard–de Hoop solution for the elastoacoustic response is given in this section. The probabilistic models associated with the three uncertain parameters are constructed in Sec. IV. Section V is devoted to the stochastic solver and to the estimation of the mean value, the variance, and the probability distribution of the random solution (variable modeling the apparent velocity of longitudinal waves). The mean values of the solid material parameters used for the computations, obtained from the literature on cortical bone, are presented in Sec. VI. The analysis of the random solution is presented in Sec. VII for a set of cortical bone samples. Finally, in Sec. VIII we outline a conclusion and in the Appendix A deals with some of the technical aspects in greater detail.

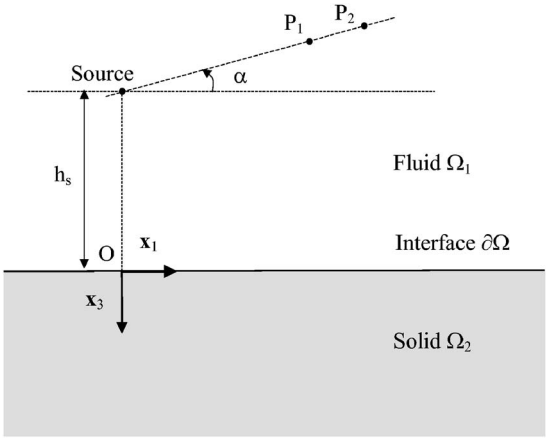


FIG. 1. Model configuration. Receivers are located at points P_1 and P_2.

II. PROBLEM FORMULATION

A. Configuration and definitions

The geometry adopted is shown in Fig. 1. It consists in the superposition of two semi-infinite media separated by a plane interface. The upper medium is a homogeneous acoustic fluid; the lower one is a homogeneous, isotropic, and linear elastic solid.

The position is specified through the Cartesian coordinates (x_1,x_2,x_3) with respect to a Cartesian reference frame $\mathbf{R}(O;\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$, where O is the origin of the space and $(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$ is an orthonormal basis of this space. The x_3 axis is chosen downward and normal to the fluid–solid interface. The fluid occupies the unbounded domain Ω_1 defined as the half-space $x_3<0$ and the solid the unbounded domain Ω_2 defined as the half-space $x_3>0$. The plane interface $\partial\Omega$ has the equation $x_3=0$.

The time coordinate is denoted by t. The fluid and the solid are at rest at times $t<0$. At $t=0$, a line source parallel to $(O;\mathbf{x}_2)$, placed in the fluid at a distance h_s from the interface, generates a cylindrical wave. Due to the nature of the source and to the geometrical configuration, the transverse waves polarized in the $(\mathbf{x}_1,\mathbf{x}_2)$ plane are not excited. The present study is conducted in the plane $(O;\mathbf{x}_1,\mathbf{x}_3)$. The total elastoacoustic wave motion will be independent of x_2, hence, all derivatives with respect to x_2 vanish in the partial differential equations that govern the wave motion. Consequently, coordinate x_2 is implicit in the mathematical expressions to follow.

The acoustic response is calculated in terms of pressure amplitudes at two receivers P_1 and P_2 of coordinates $(x_1^{(1)},x_3^{(1)})$ and $(x_1^{(2)},x_3^{(2)})$, respectively, in the plane $(O;\mathbf{x}_1,\mathbf{x}_3)$. The source P_1 and P_2 are placed on a same line. This source and receiver configuration is typical of the device used in the ultrasonic axial transmission technique (ATT) for the evaluation of the cortical layer of bone; the angle α allows us to take into account the orientation of the device with respect to the interface.

B. Equations for the fluid

The acoustic problem in the fluid is characterized by $p(\mathbf{x},t)$ and $\mathbf{v}(\mathbf{x},t)$, where $p(\mathbf{x},t)$ denotes the disturbance of the pressure from its undisturbed value at equilibrium in the reference configuration and $\mathbf{v}(\mathbf{x},t)$ is the fluid particle velocity. The components in $\mathbf{R}$ of the fluid velocity are denoted v_i.

The fluid is assumed to be homogeneous, and gravitational effects are neglected in the linear acoustic equations. In the framework of the linearized theory, it is assumed that $p(\mathbf{x},t),\mathbf{v}(\mathbf{x},t)$ and their gradients are small, but also of the same order. The constitutive equation is defined with the inverse of the fluid compressibility denoted by K and the mass density denoted by $\underline{\rho_f}$. The wave velocity in the fluid is then defined as $c_f=\sqrt{K/\rho_f}$ and the wave slowness as $s_f=1/c_f$. The equation of motion and the constitutive equation are

$$\partial_t p(\mathbf{x},t) = -\rho_f\, \partial_t v_i(\mathbf{x},t), \quad \forall\, \mathbf{x} \in \Omega_1, \tag{1}$$

$$\partial_t p(\mathbf{x},t) + K\, \partial_i v_i(\mathbf{x},t) = K\phi_V(t)\,\delta(x_1,x_3+h_s), \quad \forall\, \mathbf{x} \in \Omega_1. \tag{2}$$

The convention of summation on repeated indices is used. Derivatives with respect to t and x_i are denoted ∂_t and ∂_i, respectively, and $\delta(x_1,x_3)$ denotes the Dirac delta function where $x_1=x_3=0$. The term on the right-hand side of Eq. (2) introduces a line source of acoustic waves parallel to $(O;\mathbf{x}_2)$ characterized by its history $\partial_t\phi_V(t)$.

C. Equations for the solid (mean solid model)

The elastodynamic problem in the solid is characterized by $\boldsymbol{\sigma}(\mathbf{x},t)$ and $\mathbf{u}(\mathbf{x},t)$, where $\boldsymbol{\sigma}(\mathbf{x},t)$ denotes the stress tensor and $\mathbf{u}(\mathbf{x},t)$ denotes the particle displacement. Their components in $\mathbf{R}$ are, respectively, denoted σ_{ij} and u_i.

The solid is assumed to be linear elastic, homogeneous, and isotropic. The constitutive equation is defined with the Young modulus E, the Poisson ratio ν, and the mass density ρ_s. The wave velocities in the solid are denoted by c_L and c_T, where the letters L and T are associated with longitudinal waves and transverse waves polarized in the direction $\mathbf{x}_3$, respectively. Wave slownesses are defined as $s_L=1/c_L$ and $s_T=1/c_T$.

In Sec. IV, E,ν, and ρ_s will be modeled as random variables. In this section, the deterministic mechanical problem is presented in terms of the mean values of the random mechanical parameters. With the notation adopted, the mean value of a quantity is underlined: the mean values of random variables associated with E,ν, and ρ_s are $\underline{E},\underline{\nu}$, and $\underline{\rho_s}$, respectively. The "mean mechanical model" refers to the mechanical model for which the random parameters take their mean values.

The body force field is neglected. The equations of motion in the solid are then

$$-\partial_j\sigma_{ij}(\mathbf{x},t) + \underline{\rho_s}\,\partial_t^2 u_i(\mathbf{x},t) = 0, \quad \forall\, \mathbf{x} \in \Omega_2. \tag{3}$$

Introducing the linearized strain tensor as $\epsilon_{ij}=(1/2)(\partial_i u_j+\partial_j u_i)$, the constitutive equation (Hooke's law) is written as

$$\sigma_{ij} = c_{ijkq}\epsilon_{kq}, \tag{4}$$

where c_{ijkq} is the fourth-order stiffness tensor, which is written for an isotropic solid as

$$c_{ijkq} = \frac{\underline{E}}{(1+\underline{\nu})}\left(\frac{\underline{\nu}}{(1-2\underline{\nu})}\delta_{ij}\delta_{kq} + \frac{1}{2}(\delta_{ik}\delta_{jq}+\delta_{iq}\delta_{jk})\right),$$

where δ_{ij} is the Kronecker symbol.

In the solid, the wave velocities are defined by

$$\underline{c_L} = \sqrt{\frac{\underline{E}(1-\underline{\nu})}{(1+\underline{\nu})(1-2\underline{\nu})\underline{\rho_s}}}, \quad \underline{c_T} = \sqrt{\frac{\underline{E}}{2(1+\underline{\nu})\underline{\rho_s}}}. \tag{5}$$

D. Conditions at the plane interface

At the interface $\partial\Omega$ between the fluid and the solid, the following conditions are assumed:

$$v_3(\mathbf{x},t) = \partial_t u_3(\mathbf{x},t), \quad \forall\, \mathbf{x} \in \partial\Omega \tag{6}$$

$$\sigma_{33}(\mathbf{x},t) = -p(\mathbf{x},t), \quad \forall\, \mathbf{x} \in \partial\Omega \tag{7}$$

$$\sigma_{13}(\mathbf{x},t) = \sigma_{23}(\mathbf{x},t) = 0, \quad \forall\, \mathbf{x} \in \partial\Omega. \tag{8}$$

The continuity of the normal velocity is expressed by Eq. (6). The continuity of the pressure is depicted by Eq. (7). The condition of frictionless contact is expressed by Eq. (8).

III. SOLUTION OF THE MEAN PROBLEM

In an axial transmission ultrasonic experiment, the radio-frequency signals from the physical setup are processed following a certain procedure that yields an estimation of the velocity of longitudinal waves in the solid. This estimation is called the "apparent velocity of longitudinal waves." In the present modeling work, a procedure that mimics the actual physical procedure is followed in which simulated signals replace the actual acoustic signals. These simulated signals are solutions of the elastoacoustic problem defined by Eqs. (1)–(8).

In this section we give a brief account of the Cagniard–de Hoop method, which is used to derive an exact semianalytical expression of the problem defined by Eqs. (1)–(8). Next, the solution is presented. Finally, the procedure followed to calculate the apparent velocity of L waves with simulated signals is detailed.

A. Background on the method of solution

The Cagniard–de Hoop method has been introduced by Cagniard (1939) and modified by de Hoop (1960). It is a very efficient tool for solving and investigating time-domain

wave propagation problems in simple geometrical configurations. The Cagniard–de Hoop method is extensively described in the literature devoted to seismology [see Aki and Richard (1980); Kennett (1983); Pao and Gajewski (1977); van der Hijden (1987)].

With the Cagniard–de Hoop method, the solution is obtained as a sum of terms each associated with a physical wave contribution. The first step of the method consists in algebraic manipulations of the basic equations in the Fourier–Laplace domain dual of the space–time domain. In the second step, each wave contribution is identified in the Fourier–Laplace domain and transformed back to the space time domain with the Cagniard–de Hoop technique. In essence, the technique is a mathematical trick that avoids performing numerical integrations over the frequency and over the wave number. Finally, the explicit Green's functions in the time domain for each wave contribution are derived. The response to a specific source history is calculated using convolution formulas.

The expressions derived with the Cagniard–de Hoop method are exact solutions of the wave propagation problems. Their are valid in the near field as well as in the farfield.

B. Reflected wave closed form analytical solution

The fluid–solid configuration of interest in this paper has been investigated by de Hoop and van der Hijden (1983) and Grimal and Naili (2006). A detailed discussion of the method and analyses of the elastoacoustic problem can be found in these references. The notation used here is consistent with that used in reference Grimal and Naili (2006).

The response at a receiver in the fluid domain Ω_1 basically consists in two wave contributions: (*i*) the direct wave from the source to the receiver; (*ii*) the wave reflected at the interface. This reflected wave is itself the result of several wave phenomena, namely (*i*) a body wave in the fluid generated from specular reflection (according to Snell–Descartes law); (*ii*) lateral waves (also called "head waves"), that is, waves associated with energy propagated in the solid, close to interface, at the velocity of L or T waves and refracted back in the fluid; (*iii*) interface waves. The direct wave path, the specular reflected wave path, and the lateral wave path are sketched in Fig. 2.

In the present work we focus on the lateral wave contribution because, as explained in Sec. III C, it allows a direct assessment of the L-wave velocity in the solid. A theoretical presentation of lateral wave phenomena may be found in the monograph by Brekhovskikh (1960, p. 260); a description of the physical phenomena relative to the specific configuration investigated may be found in Bossy *et al.* (2002).

Upon the application of the Cagniard–de Hoop technique to the reflected wave contribution in the Fourier-Laplace domain, the contribution of the lateral waves can be isolated from the rest of the reflected wave contributions. This is possible because lateral wave contributions are confined in a time interval $t_l < t < t_{ff}$, where t_l denotes the arrival time of the lateral wave and t_{ff} denotes the arrival time of the body wave (specular reflection). Lateral wave contributions

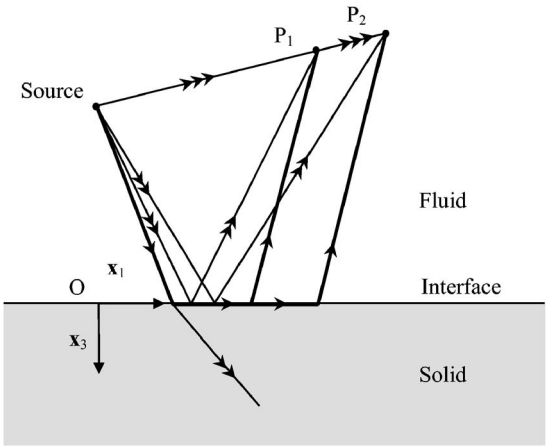


FIG. 2. Wave paths. The three types of wave paths are sketched: direct waves, reflected waves, and lateral waves (bold line).

exist only under certain conditions on the material properties of the medium. In the present work, the analysis is restricted to those cases where the lateral wave does exist and to receivers that are reached by the lateral wave. The first condition is equivalent to $\underline{c}_L > c_f$ and the second one to $|x_1|/|x_3 - h_s| > \tan\theta_c$, where θ_c is the critical angle defined by $\sin\theta_c = c_f/\underline{c}_L$, at which lateral waves are refracted and (x_1, x_3) are the coordinates of a point P in the plane $(O; \mathbf{x}_1, \mathbf{x}_3)$.

The acoustic pressure at a receiver in the fluid corresponding to the reflected waves contribution is written as

$$p_R(\mathbf{x},t) = \partial_t \phi_V(t) * \mathcal{G}(\mathbf{x},t), \tag{9}$$

where the operator * denotes the time convolution and where $\mathcal{G}(\mathbf{x},t)$, the Green's function for the reflected wave contribution, is given by Grimal and Naili (2006);

$$\begin{aligned}
&\text{if } \xi(t_{ff}) < \underline{s}_L, \quad \mathcal{G}(\mathbf{x},t) \\
&\qquad = \begin{cases} 0, & t < t_{ff}, \\ \dfrac{-\rho_f}{2\pi\sqrt{t^2 - t_{ff}^2}} \Re[R_{ff}(\xi)], & t > t_{ff}; \end{cases} \\
&\text{if } \xi(t_{ff}) > \underline{s}_L, \quad \mathcal{G}(\mathbf{x},t) \\
&\qquad = \begin{cases} 0, & t < t_l, \\ \dfrac{\rho_f}{2\pi\sqrt{t_{ff}^2 - t^2}} \Im[R_{ff}(\xi)], & t_l < t < t_{ff}, \\ \dfrac{-\rho_f}{2\pi\sqrt{t^2 - t_{ff}^2}} \Re[R_{ff}(\xi)], & t > t_{ff}. \end{cases}
\end{aligned} \tag{10}$$

In Eq. (10), the operators $\Re[\cdot]$ and $\Im[\cdot]$ are, respectively, the real and imaginary parts of the quantity between the brackets. de Hoop and van der Hijden (1983) investigated, in the general case, the reflection at a fluid–solid interface with the Cagniard–de Hoop method. They presented a solution similar to Eq. (10), although in a slightly different form due to a difference in the formulation of the basic equations. The arrival time t_{ff} of the body wave is such that

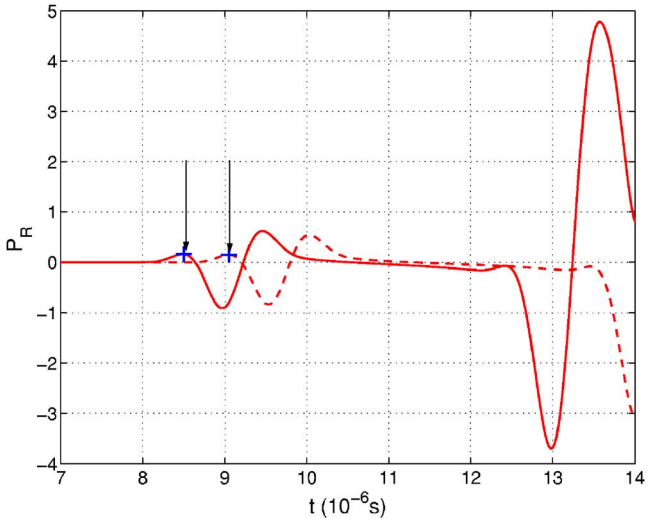


FIG. 3. (Color online) Acoustic response $p_R(\mathbf{x},t)$ *versus* times at $P_1(20,-2)$ (solid line) $P_2(22,-2)$ (dashed line), dimensions in mm. Evaluated times of flight are denoted by crosses.

$$t_{ff} = s_f r, \tag{11}$$

in which $r^2=(h_s-x_3)^2+x_1^2$ is the square of the distance between the source and the receiver. The arrival time t_l of the lateral wave contribution is defined by

$$t_l = \gamma_f(h_s - x_3) + \underline{s}_L x_1, \tag{12}$$

where $\gamma_f=\sqrt{s_f^2-\xi^2}$. The function ξ is defined on $[0,+\infty[$ and is such that

$$\xi(\tau) = \frac{\tau x_1}{r^2} \pm i\frac{h_s - x_3}{r^2}\sqrt{\tau^2 - t_{ff}^2}. \tag{13}$$

The reflection coefficient R_{ff} at the fluid–solid interface is given by

$$R_{ff} = \frac{4\underline{\mu}\gamma_f\Delta_R - \gamma_L\rho_f\underline{s}_T^2}{4\underline{\mu}\gamma_f\Delta_R + \gamma_L\rho_f\underline{s}_T^2}, \tag{14}$$

where $\gamma_L=(\underline{s}_L^2-\xi^2)^{1/2}$, $\gamma_T=(\underline{s}_T^2-\xi^2)^{1/2}$, $\chi=0.5\underline{s}_T^2-\xi^2$, and $\Delta_R=\gamma_L\gamma_T\xi^2+\chi^2$. The values taken by ξ must be such that $\Re[\gamma_L]\geqslant 0$ and $\Re[\gamma_T]\geqslant 0$. Lamé's coefficient $\underline{\mu}$ is defined by $\underline{\mu}=\underline{E}/2(1+\underline{\nu})$.

C. Apparent velocity of *L* waves

In a typical experimental setup of the axial transmission technique, the velocity of longitudinal waves is estimated based on the lateral wave propagation. Since the estimated velocity depends on some parameters of the setup, it is referred to as the "apparent" velocity of L waves, and denoted by v. In the ideal limit case where measurement errors are zero, the apparent velocity is equal to the longitudinal wave velocity.

The procedure for calculating v is the following.

(a) Pressure signals, as shown in Fig. 3, are calculated with the Cagniard–de Hoop technique with Eq. (10) at the two receivers, P_1 and P_2, separated by a distance d.

(b) The arrival time of the wave is defined as the first local maximum of the pressure that corresponds to the first zero of the function $\partial_t^2\phi_V(t)*\mathcal{G}(\mathbf{x},t)$. The time delay separating the arrival times at the two receivers is denoted by Δt.

(c) The apparent velocity of the L wave is then

$$v = \frac{d}{\Delta t}. \tag{15}$$

Wave velocity v calculated in this way is, in practice, a good approximation of the actual velocity of longitudinal waves in the solid. The apparent velocity corresponding to the mean model is denoted $\underline{v}$. Under the two following assumptions, the equality $\underline{v}=\underline{c}_L$ is verified: (*i*) the waveform remains unchanged during propagation between the two receivers; (*ii*) the line joining the receivers is parallel to the interface ($\alpha=0$). The first assumption is, in general, not strictly verified because of close field effects [Bossy (2003); Grimal and Naili (2006)]. Whether or not the second assumption is verified depends on the positioning of the ultrasonic probe with respect to the fluid–solid interface. The dependence $\underline{v}(\alpha)$ is investigated in Sec. VII.

IV. PROBABILITY MODEL OF THE UNCERTAIN PARAMETERS

Uncertainties in the transient elastoacoustics problem defined by Eqs. (1)–(4) with boundary conditions (6)–(8) are modeled using a parametric probabilistic approach. First, the uncertain parameters must be identified. Then an appropriate probabilistic model must be constructed for each uncertain parameter which is modeled by a random variable.

Since our final aim in this work is the identification of the mechanical parameters of the solid, probability models are constructed for the Young modulus E, the Poisson ratio ν and the mass density ρ_s. (In a study with a different aim, one may as well consider introducing probability models for geometrical parameters and fluid parameters.) The heterogeneity of the solid material is not addressed in this paper; the solid is assumed to have material parameters constant over space.

In order to construct a coherent probability model, only the available information on the random mechanical parameters is used. This approach avoids the use of any additional speculated information that would yield a nonphysical estimation of the probability distribution. In the context of information theory, Shannon (1948) introduced an entropy as the measure of uncertainty for probability distributions. In the context of statistical mechanics, Jaynes (1957a and 1957b) used this measure to define the maximum entropy principle for the construction of a probability distribution. This principle consists in maximizing the entropy subjected to constraints defined by the available information. The probability models constructed for the mechanical properties in the present work constitute particular cases of the ones described in Soize's works [Soize (2001, 2005)].

With the nature of the available information used for the probabilistic models, the application of the maximum entropy principle yields independent probability density functions for E, ν, and ρ_s. In other words, the independence of

the random variables associated with E, ν, and ρ_s is a consequence of the use of the maximum entropy principle.

A. Young's modulus

The Young modulus is modeled by a random variable $\mathbb{E}$ with the probability density function defined using the following information: (1) The support of the probability density function is $]0,+\infty[$. (2) The mean value is such that, by construction, $\mathcal{E}\{\mathbb{E}\}=\underline{E}$, where $\mathcal{E}$ denotes the mathematical expectation. (3) The second-order moment of its inverse is finite $\mathcal{E}\{1/\mathbb{E}^2\}<+\infty$. Information (1) is the thermodynamic restriction on Young's modulus. Information (3) is due to the ellipticity property of the random operator modeling the elasticity tensor [see Soize (2001, 2004)]. Let Y_1 be the random variable such that $\mathbb{E}=\underline{E}Y_1$, and then $\mathcal{E}\{Y_1\}=1$. The probability density function $f_{\mathbb{E}}$ of $\mathbb{E}$ is such that $f_{\mathbb{E}}(E)dE=f_{Y_1}(y_1)dy_1$. To construct the probability density function, information (3) is taken into account by requiring that $\mathcal{E}\{\ln(Y_1)\}=c_1$ with $c_1<+\infty$, where ln designates the natural logarithm function [see Soize (2001)]. To summarize, the probability density function f_{Y_1} whose support is $]0,+\infty[$ has to verify the following constraints:

$$\begin{cases} \displaystyle\int_{-\infty}^{+\infty} f_{Y_1}(y_1)dy_1 = 1, \quad \int_{-\infty}^{+\infty} y_1 f_{Y_1}(y_1)dy_1 = 1, \\ c_1 < +\infty, \quad \text{with } c_1 = \displaystyle\int_{-\infty}^{+\infty} \ln(y_1) f_{Y_1}(y_1)dy_1. \end{cases} \tag{16}$$

The application of the maximum entropy principle yields the probability density function of Y_1,

$$f_{Y_1}(y_1) = \mathbf{1}_{]0,+\infty[}(y_1)\exp(-\lambda_0 - \lambda_1 y_1) y_1^{-\lambda_2}, \tag{17}$$

where $\mathbf{1}_{]0,+\infty[}(y)$ is such that, for $B\subset\mathbb{R}$, $\mathbf{1}_B(y)=1$ if $y\in B$ and 0 if $y\notin B$, where $\mathbb{R}$ designates the set of real numbers. The three Lagrange multipliers λ_0, λ_1, and λ_2 introduced in Eq. (17) are real numbers associated with the three constraints defined by Eq. (16), and can be obtained in closed form: $\lambda_0=-\ln(\lambda_1^{\lambda_1}/\Gamma(\lambda_1))$, $\lambda_1=1/\delta_1^2$, and $\lambda_2=1-\lambda_1$, where $\delta_1^2=\mathcal{E}\{Y_1^2\}-1$ is the variance of Y_1. Parameter δ_1 is the coefficient of variation of the random variable Y_1 and can be used to control the dispersion of Y_1 and, consequently, of the random variable $\mathbb{E}$. Using the obtained expressions, Eq. (17) yields

$$f_{Y_1}(y_1) = \mathbf{1}_{]0,+\infty[}(y_1)\left(\frac{1}{\delta_1^2}\right)^{1/\delta_1^2}\frac{1}{\Gamma(\delta_1^{-2})} y_1^{(1/\delta_1^2)-1}\exp\left(-\frac{y_1}{\delta_1^2}\right), \quad \text{with } 0\leqslant\delta_1\leqslant 1/\sqrt{2}, \tag{18}$$

where $\Gamma(x)=\int_0^{+\infty} t^{x+1}x^t\,dt$ is the Gamma function.

B. Poisson's ratio

The Poisson ratio is modeled by a random variable Y_2 with the probability density function defined using the following information: (1) The support of the probability density function is $]-1,1/2[$. (2) By construction, $\mathcal{E}\{Y_2\}=\underline{\nu}$. (3) $\mathcal{E}\{(1-Y_2)^2/(1+Y_2)^2(1-2Y_2)^2\}=c_2$, with $c_2<+\infty$. Information (1) is the thermodynamic restriction on Poisson's ratio.

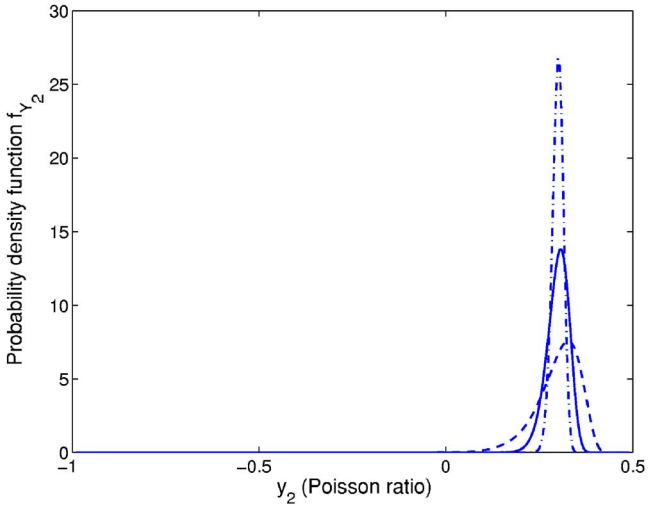


FIG. 4. (Color online) Probability density function of random variable Y_2 modeling the Poisson ratio for coefficient of variation $\delta_2=0.05$ (dash-dotted line), $\delta_2=0.1$ (solid line), and $\delta_2=0.2$ (dashed line).

Information (3) is a nontrivial condition that is required to ensure that the second-order moments of the wave velocities are finite. The derivation of this condition is detailed in Appendix A.

To summarize, the probability density function f_{Y_2} whose support is $]-1,1/2[$ has to verify the following constraints:

$$\begin{cases} \displaystyle\int_{-\infty}^{+\infty} f_{Y_2}(y_2)dy_2 = 1, \quad \int_{-\infty}^{+\infty} y_2 f_{Y_2}(y_2)dy_2 = \underline{\nu}, \\ c_2 < +\infty, \quad \text{with } c_2 = \displaystyle\int_{-\infty}^{+\infty} \frac{(1-y_2)^2}{(1+y_2)^2(1-2y_2)^2} f_{Y_2}(y_2)dy_2. \end{cases} \tag{19}$$

The application of the maximum entropy principle yields the probability density function of Y_2,

$$f_{Y_2}(y_2) = \mathbf{1}_{]-1,1/2[}(y_2) \times\exp\left(-\lambda_0-\lambda_1 y_2-\lambda_2\frac{(1-y_2)^2}{(1+y_2)^2(1-2y_2)^2}\right), \tag{20}$$

where the Lagrange multipliers λ_0, λ_1, and λ_2 cannot be found in closed form. They are obtained by minimizing the strictly convex function $H_2(\lambda_0,\lambda_1,\lambda_2)$, defined by

$$H_2(\lambda_0,\lambda_1,\lambda_2) = \lambda_0+\lambda_1\underline{\nu}+\lambda_2 c_2 + \int_{-1}^{1/2}\exp\left(-\lambda_0-\lambda_1 y_2 - \lambda_2\frac{(1-y_2)^2}{(1+y_2)^2(1-2y_2)^2}\right)dy_2. \tag{21}$$

The strictly convex optimization problem is solved numerically [Ciarlet (1989)] and c_2 is rewritten as a function of the coefficient of variation $\delta_2=\sigma_{Y_2}/\underline{\nu}$ of random variable Y_2, where σ_{Y_2} is the standard deviation given by Eq. (20). Parameter δ_2 allows the dispersion to be controlled. Figure 4 displays the probability density function f_{Y_2} for $\delta_2=0.05$ (dash-dotted line), 0.1 (solid line) and 0.2 (dashed line).

C. Mass density

The mass density is modeled by a random variable R for which the available information is the following: (1) The support of the probability density function is $]0,+\infty[$. (2) By construction, $\mathcal{E}\{\mathrm{R}\}=\underline{\rho}_s$. (3) $\mathcal{E}\{1/\mathrm{R}^2\}<+\infty$. Information (1) is the thermodynamic restriction on mass density. Information (3) is required to get a second-order solution of the stochastic problem [see Soize (2001, 2004)]. Let Y_3 be the random variable such that $\mathrm{R}=\underline{\rho}_s Y_3$, and then, $\mathcal{E}\{Y_3\}=1$. The probability density function f_R of R is such that $f_R(\rho_s)d\rho_s = f_{Y_3}(y_3)dy_3$. Information (3) is taken into account by requiring that $\mathcal{E}\{\ln(Y_3)\}=c_3$ with $c_3<+\infty$. To summarize, the probability density function f_{Y_3} whose support is $]0,+\infty[$ has to verify the following constraints:

$$\begin{cases} \displaystyle\int_{-\infty}^{+\infty} f_{Y_3}(y_3)dy_3 = 1, \quad \int_{-\infty}^{+\infty} y_3 f_{Y_3}(y_3)dy_3 = 1, \\ c_3 < +\infty, \quad \text{with } c_3 = \displaystyle\int_{-\infty}^{+\infty} \ln(y_3) f_{Y_3}(y_3)dy_3. \end{cases} \tag{22}$$

These constraints are the same as those verified by the probability density function f_{Y_1}. Using the same method as in Sec. IV A yields

$$f_{Y_3}(y_3) = \mathbf{1}_{]0,+\infty[}(y_3)\left(\frac{1}{\delta_3^2}\right)^{1/\delta_3^2}\frac{1}{\Gamma(\delta_3^{-2})}y_3^{(1/\delta_3^2)-1} \times \exp(-y3/\delta_3^2), \quad \text{with } 0 \leqslant \delta_3 \leqslant 1/\sqrt{2}, \tag{23}$$

where $\delta_3^2=\mathcal{E}\{Y_3^2\}-1$ is the variance of Y_3. The coefficient of variation δ_3 can be used to control the dispersion of the random variable Y_3 and, consequently, on the random variable R.

V. STOCHASTIC SOLVER FOR THE UNCERTAIN MECHANICAL SYSTEM

The stochastic solver used is based on a Monte Carlo numerical simulation. First, realizations of random variable $\mathbf{Y}=(Y_1,Y_2,Y_3)$ are constructed. For each realization of $\mathbf{Y}$, the corresponding realization of the random apparent velocity V is calculated. Finally, the mathematical statistics are used for constructing the estimations and a convergence analysis is performed with respect to the number of realizations.

A. Acoustic pressure

The first step consists in constructing, with the adapted generator, independent realizations $\mathbf{Y}(\theta_1),\ldots,\mathbf{Y}(\theta_n)$ using the probability density functions defined by Eqs. (18), (20), and (23). In a second step, the quantities $\mathrm{V}(\theta_1),\ldots,\mathrm{V}(\theta_n)$ are calculated using the procedure described in Sec. III with the solution, Eqs. (9)–(14), obtained with the Cagniard–de Hoop method and Eq. (15). The equations of Sec. III B are rewritten below in order to exhibit the random quantities.

For each realization $\theta_j (j=1,\ldots,n)$, the pressure at the receiver is given by

$$P_R(\mathbf{x},t,\theta_j) = \partial_t \phi_V(t) * G(\mathbf{x},t,\theta_j), \tag{24}$$

where G is the Green's function given by the following:

$$\begin{aligned} &\text{if } \xi(t_{ff}) < S_L(\theta_j), \quad \mathrm{G}(\mathbf{x},t,\theta_j) \\ &\qquad = \begin{cases} 0, & t < t_{ff}; \\ \dfrac{-\rho_f}{2\pi\sqrt{t^2-t_{ff}^2}}\Re[R_{ff}(\xi,\theta_j)], & t > t_{ff}; \end{cases} \\ &\text{if } \xi(t_{ff}) > S_L(\theta_j), \quad \mathrm{G}(\mathbf{x},t,\theta_j) \\ &\qquad = \begin{cases} 0, & t < T_l(\theta_j), \\ \dfrac{\rho_f}{2\pi\sqrt{t_{ff}^2-t^2}}\Im[R_{ff}(\xi,\theta_j)], & T_l(\theta_j) < t < t_{ff}, \\ \dfrac{-\rho_f}{2\pi\sqrt{t^2-t_{ff}^2}}\Re[R_{ff}(\xi,\theta_j)] & t > t_{ff}; \end{cases} \end{aligned} \tag{25}$$

where $S_L=1/C_L$ and $S_T=1/C_T$ are the random variables modeling the wave slownesses, respectively, associated with s_L and s_T. The arrival time of the body wave t_{ff} is defined by Eq. (11). In contrast, in the probabilistic model, the arrival time of the lateral wave contribution is modeled by the random variable T_l, defined by

$$T_l = \gamma_f(h_s - x_3) + S_L x_1. \tag{26}$$

The function ξ is given by Eq. (13). The realization $R_{ff}(\xi,\theta_j)$ of the random reflection coefficient associated with $R_{ff}(\xi)$ is written as

$$R_{ff}(\xi,\theta_j) = \frac{4\mu_j\gamma_f\Delta_{R,j} - \gamma_{L,j}\rho_f S_{T,j}^2}{4\mu_j\gamma_f\Delta_{R,j} + \gamma_{L,j}\rho_f S_{T,j}^2}, \tag{27}$$

where $\gamma_{L,j}=(S_L^2(\theta_j)-\xi^2)^{1/2}$, $\gamma_{T,j}=(S_T^2(\theta_j)-\xi^2)^{1/2}$, $\chi_j = 0.5S_T^2(\theta_j)-\xi^2$, and $\Delta_{R,j}=\gamma_{L,j}\gamma_{T,j}\xi^2+\chi_j^2$. Realization j of the random variable associated with μ is defined by $\mu_j = \underline{E}Y_1(\theta_j)/2[1+Y_2(\theta_j)]$.

B. Convergence analysis

The convergence analysis with respect to n is carried out in studying the convergence of the estimated second-order moment of V, defined by $m_{V,2}=\mathcal{E}\{V^2\}$. An estimation of $m_{V,2}$ is provided by $m_{V,2}\simeq \mathrm{Conv}^2(n)$, where

$$\mathrm{Conv}^2(n) = \frac{1}{n}\sum_{j=1}^{n} V(\theta_j)^2. \tag{28}$$

Graphs of functions $n\rightarrow\log_{10}[\mathrm{Conv}^2(n)]$ for different values of the coefficients of variation $(\delta_1,\delta_2,\delta_3)$ are shown in Fig. 5, where $\log_{10}$ is the base 10 logarithm function. The thin solid line represents the case for which $\delta_1=\delta_2=\delta_3=0.05$. The medium solid line, the case for which $\delta_1=\delta_2=\delta_3=0.1$. And the thick solid line, the case for which $\delta_1=\delta_2=\delta_3=0.2$. Convergence is reached for $n>1000$, $n>2000$, and $n>3500$ in the three cases, respectively.

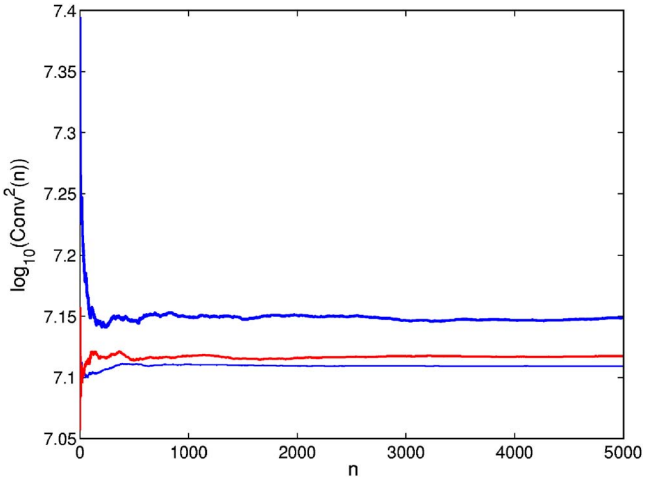


FIG. 5. (Color online) Convergence of the estimated second-order moment of V with respect to n. The coefficients of variation $\delta_i (i=1,2,3)$ are all set to the same value: $\delta_i=0.05$ (thin solid line), $\delta_i=0.1$ (medium solid line); $\delta_i=0.2$ (thick solid line).

C. Estimation of the mean value and of the confidence region

Let $v_1=V(\theta_1),\ldots,v_n=V(\theta n)$ be the n calculated independent realizations of the random variable V. Estimations of mathematical expectation $\mathcal{E}\{V\}$ and coefficient of variation δ_V are given by

$$\mathcal{E}\{V\} \simeq \underline{\hat{V}} = \frac{1}{n}\sum_{j=1}^{n} v_j \quad \text{and}$$
$$\delta_V \simeq \frac{n}{\sqrt{n-1}} \frac{\sqrt{\sum_{j=1}^{n}\left(v_j-(1/n)\sum_{k=1}^{n} v_k\right)^2}}{\sum_{j=1}^{n} v_j}. \tag{29}$$

The quantile method is used to construct the confidence region associated with a probability level P_c for random variable V. The confidence region is limited by a lower and an upper envelope denoted v^- and v^+, respectively,

$$\mathcal{P}(v^- < V \leqslant v^+) = P_c. \tag{30}$$

Let F_V be the cumulative distribution function (continuous from the right) of V defined by $F_V(v)=\mathcal{P}(V\leqslant v)$. For $0<p<1$, the pth quantile (or fractile) of F_V is defined as

$$\zeta(p) = \inf\{v:F_V(v) \geqslant p\}. \tag{31}$$

The lower and the upper envelopes for the symmetric interval are defined by

$$v^- = \zeta\left(\frac{1-P_c}{2}\right), \quad v^+ = \zeta\left(\frac{1+P_c}{2}\right). \tag{32}$$

The estimations of v^- and v^+ are performed by using the sample quantiles [Serfling (1980)]. Let $\tilde{v}_1<\cdots<\tilde{v}_n$ be the order statistics associated with $v_1,\ldots,v_n$. Therefore, we have the following estimations:

$$v^- \simeq \tilde{v}_{j^-}, \quad \text{with } j^- = \text{fix}[n(1-P_c)/2], \tag{33}$$

$$v^+ \simeq \tilde{v}_{j^+}, \quad \text{with } j^+ = \text{fix}[n(1+P_c)/2], \tag{34}$$

in which fix(z) is the integer part of real number z.

VI. CALCULATION PARAMETERS AND VALIDATION OF THE MEAN MODEL

In this section we presents the parameters used for the calculation (source history and numerical data used for the mean model) and some calculations used for the validation of the mean model.

A. A set of cortical bone samples

Dong and Guo (2004) performed mechanical tests and measured the porosity of a set of 18 cortical bone samples obtained from 6 human femurs. The authors introduced the usual assumption that cortical bone can be viewed as a homogeneous, linear elastic, and transversely isotropic material with the plane of isotropy perpendicular to the long axis of the bone. Accordingly, five elastic parameters were measured. Eighteen values were obtained for each measured parameter. Then Dong and Guo (2004) performed mathematical statistics to calculate the mean value and the standard deviation of each parameter.

In the present work, bone is modeled as an isotropic solid. Since the work focuses on the measurement of the longitudinal wave velocity along the axis of bone, only $E=E_1$, where E_1 is the Young modulus corresponding to solicitation of the bone along this axis is used in the present work. Based on their experimental values, Dong and Guo (2004) give a mean value of $\underline{E}_1=16.61\times10^9$ Pa and a value of the standard deviation, for the set of 18 samples, of $\pm1.83\times10^9$ Pa.

The mean value of the porosity ϕ for the set of samples is $\underline{\phi}=8.95\%$ with a standard deviation of $\pm4.16\%$. For the purposes of the present work, the density has been calculated based on the given values of porosity by using the formula

$$\rho_s = \rho_t(1-\phi), \tag{35}$$

where ρ_t is the mass density of the bone tissue (around the pores). The bone tissue mass density has been calculated from the data published by Raum *et al.* (2005), $\rho_t=1891$ kg m^{-3}. In the present work, ρ_t is considered to be a "universal" value valid for every bone sample; as a consequence, ρ_t is modeled as deterministic. The mean value of the mass density is then $\underline{\rho}_s=(1-\underline{\phi})\rho_t=1722$ kg m^{-3} and the standard deviation for the set of 18 samples is ±78.7 kg m^{-3}.

B. Acoustic source

The history of the acoustic source is defined by the function

$$\partial_t\phi_V(t) = \sin(2\pi tf)\exp[-4(tf-1)^2], \tag{36}$$

where f is the center frequency of the pulse. In this paper, calculations are performed with $f=1$ MHz. Figure 6 displays the graph of the function $t\to\partial_t\phi_V(t)$.

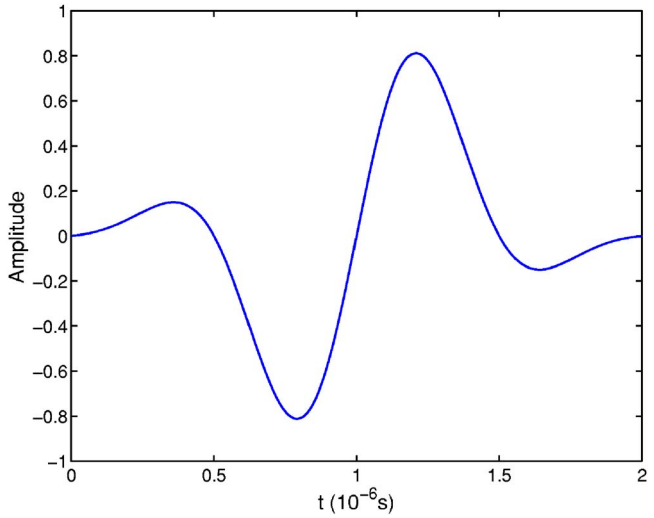


FIG. 6. (Color online) History of acoustic source $\partial_t \phi_V(t)$ with 1 MHz center frequency.

C. Data for the mean model

The source is located at $h_s=2$ mm from the interface. The distance between the source and the first receiver is 20 mm. The distance between the receivers is 2 mm. The fluid is supposed to be water: $\rho_f=1000$ kg m^{-3} and $c_f=1490$ m s^{-1}.

The mean values of the solid (bone) parameters are adapted from the experimental results detailed in Sec. VI A. Only the mean values of E and ρ obtained from the data of Dong and Guo (2004) have been used as input in the probability density functions of the random variables. In particular, the standard deviations derived from the experimental measurements were not used. Indeed, the standard deviation is not taken to be an available information for the construction of the probabilistic models with the maximum entropy principle. This is consistent with the fact that the number of bone samples measured in the experiments of Dong and Guo is too small to provide converged second-order moments (standard deviations) of the measured parameters.

The mean model is defined with $\underline{E}=16.61\times10^9$ Pa and $\underline{\rho_s}=1722$ kg m^{-3}. The mean value of the isotropic Poisson's ratio is taken to be $\underline{\nu}=0.3$, which is the value usually assumed in bone [Zysset *et al.* (1999)]. With these values for the mean model, $\underline{c_L}=3604$ m s^{-1}.

D. Validation

A first validation of the mean model has been obtained by comparing the calculated velocities with velocity measurements on several materials reported in the thesis of Bossy (2003), see pp. 69–71, in an axial transmission experimental setup, for different distances between the source and the receivers. A second validation of the mean model was obtained by comparing the calculated velocities for several values of the angle α with velocities calculated by Bossy (2003), see pp. 107–110, in the same geometrical configuration, with the help of a finite differences code based on Virieux's numerical scheme [Virieux (1986)].

TABLE I. An estimation of mathematical expectation $\mathcal{E}\{V\}$, standard deviation σ_V, and the coefficient of variation δ_V of the apparent wave velocity V.

$\underline{v}$(m s^{-1})	3574	3574	3574	3574	3574	3574	3574	3574	3574
δ_1	0.05	0.1	0.2	0	0	0	0	0	0
δ_2	0	0	0	0.05	0.1	0.2	0	0	0
δ_3	0	0	0	0	0	0	0.05	0.1	0.2
$\mathcal{E}\{V\}\simeq\underline{\hat{V}}$(m s^{-1})	3573	3570	3551	3580	3600	3670	3578	3586	3632
σ_V(m s^{-1})	89	179	347	73	147	301	90	181	363
$\delta_V(10^{-2})$	2.49	5.01	9.77	2.04	4.08	8.21	2.51	5.05	9.99

VII. RESULTS

A. Mean value and dispersion of the apparent velocity

In this section we focus on the case $\alpha=0$. The sensitivity of the apparent velocity v to each of the mechanical parameters (E, ν, and ρ_s) is presented. Each of the coefficients of variation $\delta_i(i=1,2,3)$, associated with E, ν, and ρ_s, respectively, varies separately. For this purpose, three cases are defined: Case 1 ($\delta_1=\delta, \delta_2=\delta_3=0$); case 2 ($\delta_2=\delta, \delta_1=\delta_3=0$); and case 3 ($\delta_3=\delta, \delta_1=\delta_2=0$), where δ takes the values 0.05, 0.1, and 0.2 successively.

With the values defined for the mean model, $\underline{v}=3574$ m s^{-1}. The difference between $\underline{c_L}=3604$ m s^{-1} and $\underline{v}$ is due to the changes in the shape of the response between the two receivers.

The estimated mean values $\underline{\hat{V}}$ for cases 1–3 are collected in Table I. Note that (*i*) $\underline{\hat{V}}<\underline{v}$ in case 1, while $\underline{\hat{V}}>\underline{v}$ in cases 2 and 3; (*ii*) the values of $\underline{\hat{V}}$ decrease with dispersion in case 1 while they increase in cases 2 and 3; (*iii*) the value of $\underline{\hat{V}}$ is more sensitive to the dispersion on $\nu(\delta_2)$ and $\rho_s(\delta_3)$ than on the dispersion on $E(\delta_1)$.

The confidence regions of the calculated apparent wave velocity for $P_c=0.95$ are plotted in Figs. 7–9 for cases 1–3. The width of the confidence regions are 345, 684, and 1348 m s^{-1} in case 1 for the three dispersion levels, respec-

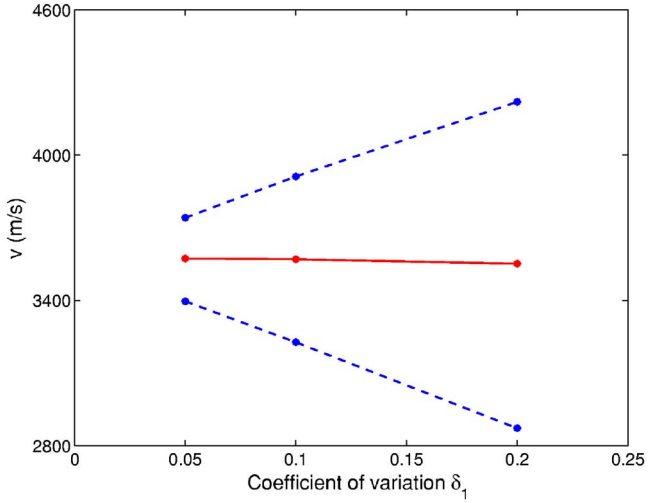


FIG. 7. (Color online) Confidence region for a probability level of $P_c=0.95$ (dashed line) and estimated mean value (solid line) of the random variable V *versus* the coefficient of variation associated with the Young's modulus.

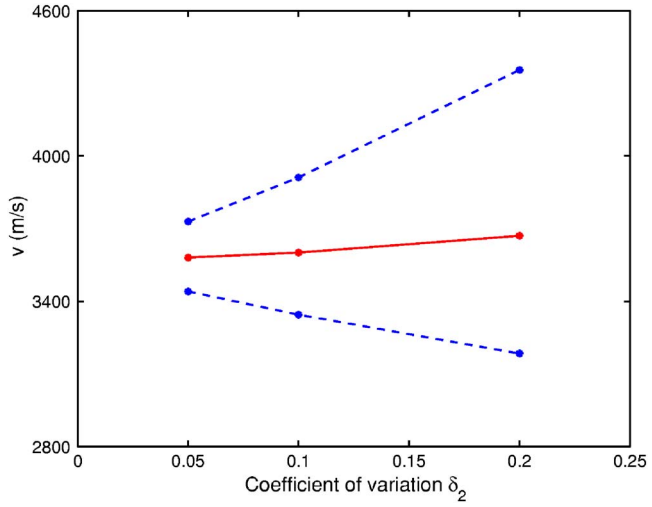


FIG. 8. (Color online) Confidence region for a probability level of P_c =0.95 (dashed line) and estimated mean value (solid line) of the random variable V *versus* the coefficient of variation associated with the Poisson ratio.

tively; 288,566, and 1160 m s^{-1} in case 2; 347,716, and 1416 m s^{-1} in case 3. The width of the confidence regions are of the same order of magnitude, however, the values indicate that the apparent velocity is slightly more sensitive to variations of E and ρ_s than to variations of ν. One may also note that the upper and lower envelopes are not symmetric with respect to mean values.

B. Sensitivity analysis with respect to angle α

The evolution of the distribution of V with respect to α and the dispersion levels on the random mechanical parameters are shown in Fig. 10. Angle α varies in $[-5,5]$ [values given in degrees (deg)] with a sampling angle step 0.5 deg. The upper and lower envelopes of the confidence region (P_c=0.95) of V *vs* α are represented by dash-dotted lines in case $\delta_1=\delta_2=\delta_3=0.05$, by thick solid lines in case $\delta_1=\delta_2=\delta_3=0.1$, and by dashed lines in case $\delta_1=\delta_2=\delta_3=0.2$. The thin solid line represents the apparent velocity $\underline{v}$ obtained from the mean model. In Fig. 10, $\underline{v}$ was plotted in order to allow a comparison between the random modeling of the axial transmission experiment and a deterministic modeling (model with the values of the mean model). Both $\underline{v}$ and the widths of the confidence regions are decreasing functions of α. In essence, this is due to the evolution with α of the wave paths of the waves arriving at the two receivers. In addition, the upper and lower envelopes are not symmetric with respect to $\underline{v}$. The ratio of the width of the confidence region for a given set of coefficients of variation and $\underline{v}$ depends weakly on α. In other words, the orientation of the receivers line has a weak influence on the normalized confidence region.

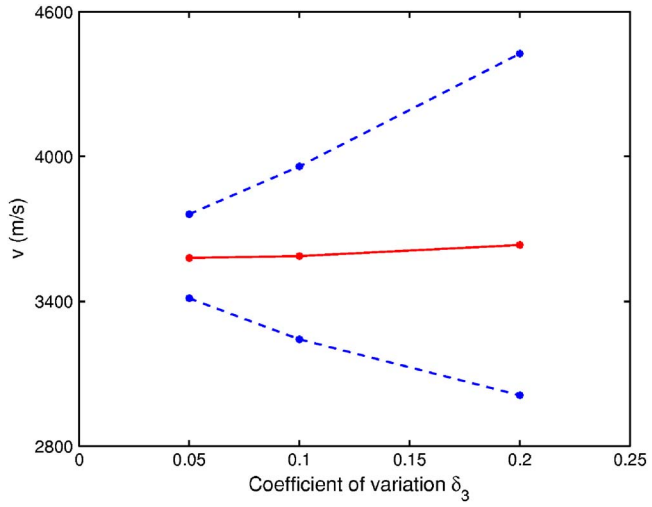


FIG. 9. (Color online) Confidence region for a probability level of P_c =0.95 (dashed line) and estimated mean value (solid line) of the random variable V *versus* the coefficient of variation associated with the mass density.

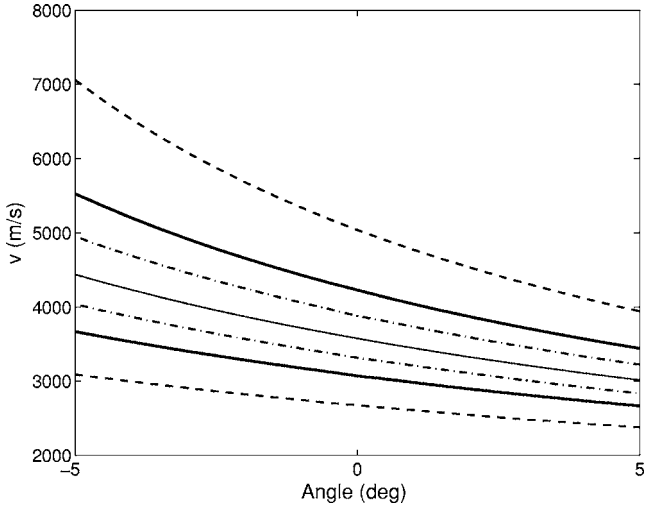


FIG. 10. Confidence region of random variable V *versus* angle α for a fixed level of probability P_c=0.95. Each plot (i.e., two similar lines) is associated with a single value of the coefficients of variation of the uncertain parameters $\delta_i (i=1,2,3)$; δ_i=0.05 (dash-dotted line), δ_i=0.1 (thick solid line) δ_i =0.2 (dashed line). The apparent velocity of longitudinal waves $\underline{v}$ obtained for the mean model *versus* angle α is plotted in thin solid line.

VIII. CONCLUSION

To our knowledge, in this paper we present the first probabilistic model of the ultrasonic axial transmission technique. This technique can measure the longitudinal wave velocity of an immersed solid, based on the propagation of a lateral wave. This work did not focus on the velocity of longitudinal waves itself but rather on an apparent velocity of longitudinal waves, as measured by existing devices. This was done in order to include some measurement constraints in the model. Measurement constraints were modeled as deterministic parameters while the material parameters of the solid were modeled as random homogeneous parameters. The reason for this choice is that we are interested in measuring solids with unknown mechanical parameters while the measurement errors are supposed to be well controlled.

The developed random model of the axial transmission technique was used to predict the apparent velocity when the solid is cortical bone. The probability density functions constructed for the random Young's modulus and the random mass density are such that their mean values are equal to mean values calculated for a set of 18 bone samples with

mathematical statistics. The values of the coefficients of variation δ_i used in the simulations for the random mechanical parameters are of the same order of magnitude as the physiological dispersion levels for cortical bone. However, the validity of the many assumptions must be tested in future work, in particular, the assumptions concerning bone isotropy.

The analysis has revealed a complex behavior of the apparent velocity with respect to the levels of dispersion of the bone material parameters. In particular, (*i*) the estimated mean value of the apparent velocity varies with the coefficients of variation, and its behavior is different with respect to these coefficients on each mechanical parameter; (*ii*) the apparent velocity is more sensitive to variations of Young's modulus or mass density than to variations of Poisson's ratio; (*iii*) the angle α between the fluid–solid interface and the source-receiver line has a strong influence on the mean value and on the width of the confidence region.

APPENDIX A: CONDITIONS ON THE WAVE VELOCITIES

Because our aim in this work is to model a technique that allows an estimation of the longitudinal wave velocity to be constructed, particular attention is paid to the moments of the velocities. At least, the second-order moment of C_L and C_T must be finite.

(*a*) *Longitudinal wave velocity*. The condition is written as

$$\exists q \in \mathbb{N}^*, \quad \text{such that } \mathcal{E}\{C_L^{2q}\} < +\infty, \tag{A1}$$

where $\mathbb{N}^*$ is the set of nonzero integers.

Using the expression of C_L deduced from Eq. (5) yields

$$\exists q \in \mathbb{N}^*, \quad \text{such that } \mathcal{E}\{C_L^{2q}\} = \mathcal{E}\left\{ \frac{\mathbb{E}^q (1-Y_2)^q}{(1+Y_2)^q (1-2Y_2)^q \mathrm{R}^q} \right\}. \tag{A2}$$

Since random variables $\mathbb{E}, Y_2$, and R are mutually independent, then

$$\mathcal{E}\{C_L^{2q}\} = \mathcal{E}\{\mathbb{E}^q\} \mathcal{E}\left\{ \frac{1}{\mathrm{R}^q} \right\} \mathcal{E}\left\{ \frac{(1-Y_2)^q}{(1+Y_2)^q (1-2Y_2)^q} \right\}. \tag{A3}$$

By construction, random variable $\mathbb{E}$ has a finite second-order moment (Sec. IV A). In Sec. IV B, the third information on the Poisson ratio is $\mathcal{E}\{(1-Y_2)^2/(1+Y_2)^2(1-2Y_2)^2\} < +\infty$. Finally, in Sec. IV C, the third information on the mass density is $\mathcal{E}\{1/\mathrm{R}^2\} < +\infty$. Then Eq. (A1) holds for $q=2$. The finite fourth-order moment of longitudinal wave velocity C_L implies a finite second-order moment of C_L.

(*b*) *Transverse wave velocity*. In order to use the third information on Y_2, we write

$$\frac{(1-Y_2)^2}{(1+Y_2)^2(1-2Y_2)^2} = \left(\frac{2}{3(1+Y_2)} + \frac{1}{3(1-2Y_2)} \right)^2. \tag{A4}$$

The mathematical expectation of the left hand of Eq. (A4) is finite (see Sec. IV B). Taking the mathematical expectation of Eq. (A4) yields

$$\mathcal{E}\left\{ \frac{(1-Y_2)^2}{(1+Y_2)^2(1-2Y_2)^2} \right\} = \mathcal{E}\left\{ \frac{4}{9(1+Y_2)^2} \right\} + \mathcal{E}\left\{ \frac{4}{9(1+Y_2)(1-2Y_2)} \right\} + \mathcal{E}\left\{ \frac{1}{9(1-2Y_2)^2} \right\}. \tag{A5}$$

Because each term on the right hand of Eq. (A5) is strictly positive and finite,

$$\mathcal{E}\left\{ \frac{1}{(1+Y_2)^2} \right\} < +\infty. \tag{A6}$$

Since the random variable C_T modeling the transverse wave velocity is such that

$$C_T^4 = \frac{\mathbb{E}^2}{4(1+Y_2)^2 \mathrm{R}^2}, \tag{A7}$$

then the fourth-order moment of C_T is finite and, consequently, its second-order moment is finite.

Aki, K., and Richard, P. (**1980**). *Quantitative Seismology: Theory and Methods* (Freeman, San Francisco).

Bossy, E. (2003). "Évaluation ultrasonore de l'os cortical par transmission axiale: modélisation et expérimentation *in vitro* et *in vivo*," Ph.D. thesis, Université Pierre et Marie Curie.

Bossy, E., Talmant, M., Defontaine, M., Patat, F., and Laugier, P. (**2004**). "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., and Laugier, P. (**2002**). "Effect of bone cortical thickness on velocity measurement using ultrasonic axial transmission: a 2D simulation study," J. Acoust. Soc. Am. **112**, 297–307.

Brekhovskikh, L. (**1960**). *Waves in Layered Media* (Academic, New York).

Cagniard, L. (**1939**). *Réflexion et Réfraction des Ondes Séismiques Progressives* [translated and revised by E. A. Flinn and C. H. Dix, Reflection and Refraction of Progressive Seismic Waves (Mc-Graw Hill, New York, 1962)] (Gauthier-Villars, Paris).

Ciarlet, P. (**1989**). *Introduction to Numerical Linear Algebra and Optimisation* (Cambridge University Press, Cambridge).

de Hoop, A. (**1960**). "A modification of Cagniard's method for solving seismic pulse problems," Appl. Sci. Res. **8**, 349–356.

de Hoop, A., and van der Hijden, J. (**1983**). "Generation of acoustic waves by an impulsive line source in a fluid/solid configuration with a plane boundary," J. Acoust. Soc. Am. **74**, 333–342.

Dong, X., and Guo, X. (**2004**). "The dependence of transversally isotropic elasticity of human femoral cortical bone on porosity," J. Biomech. **37**, 1281–1287.

Foldes, A., Rimon, A., Keinan, D., and Popovtzer, M. (**1995**). "Quantitative ultrasound of the tibia: A novel approach for assessment of bone status," Bone (N.Y.) **17**, 363–367.

Grimal, Q., and Naili, S. (in press). "A theoretical analysis in the time-domain of wave reflection on a bone plate," J. Sound Vib..

Jaynes, E. (**1957a**). "Information theory and statistical mechanics," Phys. Rev. **106** (4), 620–630.

Jaynes, E. (**1957b**). "Information theory and statistical mechanics. II," Phys. Rev. **108**, 171–190.

Kapur, J., and Kesavan, H. (**1992**). *Entropy Optimization Principles with Applications* (Academic, New York).

Kennett, B. (**1983**). *Seismic Wave Propagation in Stratified Media* (Cambridge University Press, Cambridge).

Lowet, G., and Van der Perre, G. (**1996**). "Ultrasound velocity measurements in long bones: measurement method and simulation of ultrasound wave propagation," J. Biomech. **29**, 1255–1262.

Pao, Y., and Gajewski, R. (**1977**). "The generalized ray theory and transient response of layered elastic solids," in *Physical Acoustics*, edited by W. Mason and R. Thurston (Academic, New York), Vol. **XIII**, pp. 183–265.

Raum, K., Cleveland, R., Peyrin, F., and Laugier, P. (**2005**). "Derivation of

elastic stiffness from site-matched mineral density and acoustic impedance maps," Phys. Med. Biol.. **50**, 1–12
Serfling, R. (**1980**). *Approximation Theorems of Mathematical Statistics* (Wiley, New York).
Shannon, C. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423.
Soize, C. (**2001**). "Maximum entropy approach for modeling random uncertainties in transient elastodynamics," J. Acoust. Soc. Am. **109**, 1979–1996.
Soize, C. (**2004**). "Random-field model for the elasticity tensor of anisotropic random media," C. R. Mec. **332**, 1007–1012.
Soize, C. (**2005**). "Random matrix theory for modeling uncertainties in computational mechanics," Comput. Methods Appl. Mech. Eng. **194**, 1333–1366.
van der Hijden, J. (**1987**). *Propagation of transient elastic waves in stratified anisotropic media*, Vol. **32** of Applied Mathematics and Mechanics (North-Holland, Elsevier Science Publishers, Amsterdam).
Virieux, J. (**1986**). "*P-SV* Wave propagation in heterogeneous media: velocity-stress finite-difference method," Geophysics **51**, 889–901.
Zysset, P., Guo, X., Hoffler, C., Moore, K., and Goldstein, S. (**1999**). "Elastic modulus and hardness of cortical and trabecular bone lamellae measured by nanoindentation in the human femur," J. Biomech. **32**, 1005–1012.

Two-dimensional problems of diffraction by finite collinear structures

Bair V. Budaev[a] and David B. Bogy[b]
Department of Mechanical Engineering, University of California, Berkeley, California 94720

(Received 7 October 2005; revised 28 November 2005; accepted 28 November 2005)

Two closely related problems of diffraction are solved by use of the probabilistic random walk method. The first concerns diffraction by a boundary of a half-space with a piecewise constant boundary impedance, and the second solves the problem of diffraction by a finite segment with different impedances on its sides. The solutions are represented as superpositions of predefined geometric fields with several diffracted fields, which are explicitly represented as mathematical expectations of certain functionals along the trajectories of specified random motions running across multisheet analytic manifolds associated with the boundary conditions. Numerical examples confirm the feasibility of the solutions. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159292]

PACS number(s): 43.20.Bi, 43.20.El, 43.20.Fn, 43.20.Dk [RMW] Pages: 741–750

I. INTRODUCTION

Problems arising in the study of wave interactions with bounded objects may be extremely complex even in the simplest geometrical configurations where one may not expect difficulties. For example, assume a plane wave $u_*=e^{-ik(x\cos\theta_*+y\sin\theta_*)}$ illuminates an ideally reflecting boundary, $y=0$, of the two-dimensional half-space $y>0$. Then, the total field generated in the half-plane consists of the incident wave u_* and the reflected wave $u_r=Ke^{-ik(x\cos\theta_*+y\sin\theta_*)}$, where $K=-1$ in the case when the total field satisfies the Dirichlet boundary conditions, or $K=1$ in the case of Neumann conditions. This simple description of the reflected waves is easily extendable to the more general impedance boundary condition $(u_y'+\mathrm{i}Bu)|_y=0$ with a constant impedance B, and it is more difficult, although still possible, to describe the wave fields analytically in the case when the impedance has one constant value in the domain $x<a$ and another constant value in the domain $x>a$, where a is an arbitrary point. However, if the boundary of the half plane is subdivided into three parts with different, but not necessarily distinct, constant impedances the problem of computation of the total wave field becomes quite complex, and it does not have a conventional closed-form solution.

Most recently, the problem of diffraction by a three-part impedance plane was addressed in Ref. 1 on the basis of the Sommerfeld-Maliuzhinets spectral representation of the wave fields. In that paper the problem is reduced to a system of functional equations which are then converted to independent integral equations. In the case when the central piece of the three-part boundary is short compared to the wavelength, the obtained integral equations are solved asymptotically. In the past, similar problems have also been treated by the method of integral equations or by the Wiener-Hopf method, which also leads to integral equations. A brief review of such work is presented in Ref. 1, and so we do not elaborate on it here in more detail.

In this paper we explicitly solve two closely related two-dimensional problems of diffraction by finite collinear structures: (i) the problem of diffraction by the boundary of a half-space with an $(N+1)$-piecewise constant boundary impedance, and (ii) the problem of diffraction by a finite-length segment with different impedances on its sides. The solutions are represented as superpositions of the geometric fields that are predetermined by geometrical-optics with a number of diffracted fields $U_1, U_2, \ldots$, centered at the points $O_1, O_2, \ldots$, where the impedance has jumps. Every diffracted field is represented as the product $U_n=e^{\mathrm{i}kr_n}u_n(r_n,\theta_n)$, where (r_n,θ_n) are polar coordinates centered at the points O_n. The amplitudes $u_n(r_n,\theta_n)$ are determined from the complete transport equations $\nabla^2 u_n+2\mathrm{i}k\,\partial u_n/\partial\theta_n+\mathrm{i}ku_n/r_n=0$, which are solved by explicit Feynman-Kac formulas as mathematical expectations of specified functionals on trajectories of random motions determined by the geometry of the problem. This approach is based on the random walk method developed in Refs. 2 and 3, but here we employ random motions running across a specific multisheet analytic manifold associated with the boundary conditions.

These two problems are treated in the following two sections. Section II is devoted to the problem of diffraction in a half-plane with a piecewise continuous boundary impedance, and Sec. III addresses the problem of diffraction by a finite segment. Each of the sections ends with numerical examples illustrating several cases of the obtained solutions, and Sec. II provides a brief introduction to the stochastic processes employed in this paper.

II. DIFFRACTION IN A HALF PLANE WITH A PIECEWISE CONSTANT BOUNDARY IMPEDANCE

A. Formulation of the problem

The problem of diffraction of a plane wave

$$U_*=e^{-ik(x\cos\theta_*+y\sin\theta_*)},\quad 0<\theta_*<\pi, \tag{2.1}$$

in a half-plane $y>0$ with the boundary impedance $B(x)$ consists of the computation of the bounded solution of the Helmholtz equation

[a]Electronic mail: budaev@berkeley.edu
[b]Electronic mail: dbogy@cml.me.berkeley.edu

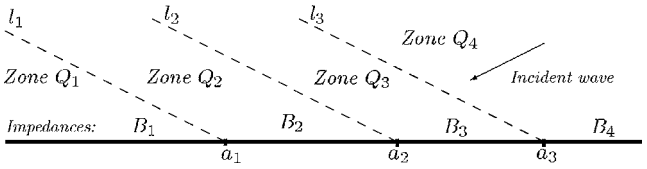


FIG. 1. Diffraction in a half-plane with a piecewise constant boundary impedance for the particular case $N=3$.

$$\nabla^2 U + k^2 U = 0, \tag{2.2}$$

which satisfies the impedance boundary conditions

$$\left.\frac{\partial U(x,y)}{\partial y} + ikB(x)U(x,y)\right|_{y=0} = 0, \tag{2.3}$$

and which does not contain waves arriving from infinity except for the incident wave U_*. If the generally complex coefficient $B(x)$ is restricted by

$$\mathrm{Re}(B) \geqslant 0, \tag{2.4}$$

then the problem is well-posed, but the actual computation of its solution is not a trivial task, except for the trivial case $B(x)=\text{const}$, and a somewhat more complicated case where

$$B(x) = \begin{cases} B_1 = \text{const} & \text{if } x < a \\ B_2 = \text{const} & \text{if } x > a, \end{cases} \tag{2.5}$$

which also admits an explicit closed-form solution.[4]

Here we study the more general problem of a half-plane with an arbitrary piecewise constant boundary impedance

$$B(x) = B_n = \text{const} \quad \text{on } x \in I_n, \quad 1 \leqslant n \leqslant N+1, \tag{2.6}$$

where B_n are constants restricted by Eq. (2.4), and I_n are defined as the intervals

$$I_1 = (-\infty, a_1),\ I_2 = (a_1, a_2),\ \ldots,\ I_{N+1} = (a_N, \infty), \tag{2.7}$$

separated by N distinct points $a_1 < a_2 < \cdots < a_N$, as shown in Fig. 1.

Elementary geometric-optical analysis suggests that the solution of the problem with N discontinuities of the impedance can be sought as a superposition

$$U = U_* + U_r + U_1 + U_2 + \cdots + U_N, \tag{2.8}$$

where the first two terms U_* and U_r are defined in advance, while the other N terms $U_1, U_2, \ldots, U_N$, which will be referred to hereafter as diffracted waves, are yet to be determined. The first term U_* is the incident wave (2.1). The second term is a piecewise continuous reflected wave

$$U_r = R_m e^{-ik(x\cos\theta_* - y\sin\theta_*)} \quad \text{if } (x,y) \in Q_m, \tag{2.9}$$

where the reflection coefficients

$$R_m = \frac{\sin\theta_* - B_m}{\sin\theta_* + B_m}, \quad 1 \leqslant m \leqslant N+1, \tag{2.10}$$

are associated with the domains

$$Q_m: \quad x + y\cot\theta_* \in I_m, \quad 1 \leqslant m \leqslant N+1, \tag{2.11}$$

separated by the lines

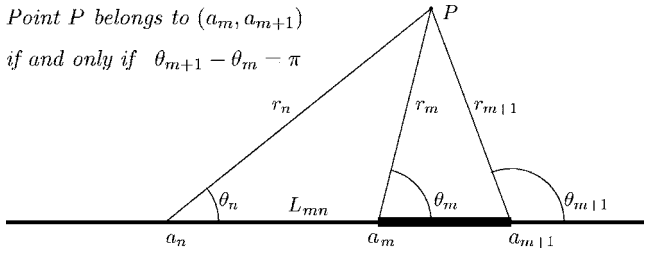


FIG. 2. Polar coordinates (r_n, θ_n) and (r_m, θ_m).

$$l_n: \quad (x - a_n)\sin\theta_* + y\cos\theta_* = 0, \quad 1 \leqslant n \leqslant N, \tag{2.12}$$

shown in Fig. 1 for the particular case when $0 < \theta_* < \pi/2$ and $N=3$.

It is obvious that the sum $(U_* + U_r)$ satisfies all of the conditions of the problem of diffraction except that it is discontinuous along each of the lines l_n defined in Eq. (2.12). Correspondingly, we require that every diffracted field U_n obeys the Helmholtz equation (2.2) and the boundary conditions (2.3). Additionally, to compensate the discontinuity of $(U_* + U_r)$ along the nth line l_n, we require that the nth field U_n obey the interface conditions

$$U_n\big|_{l_n^-} - U_n\big|_{l_n^+} = (R_n - R_{n+1})\exp\left(-ik\frac{x - a_n}{\cos\theta_*} - ika_n\cos\theta_*\right), \tag{2.13}$$

$$\left.\frac{\partial U_n}{\partial n}\right|_{l_n^-} = \left.\frac{\partial U_n}{\partial n}\right|_{l_n^+}, \tag{2.14}$$

where l_n^- and l_n^+ are the left and the right sides of l_n, respectively.

B. Computation of the diffracted fields

To describe the diffracted fields U_n with $1 \leqslant n \leqslant N$ it is convenient to introduce N different polar coordinates (r_n, θ_n) centered at the points with the Cartesian coordinates $x=a_n$, $y=0$.

The main advantage of the new coordinate systems is that they provide analytic criteria for determining whether or not a given point P belongs to a chosen interval I_m between a_m and a_{m+1}. To formulate such criteria we apply the cosine theorem to the triangle $a_m P a_{m+1}$ shown in Fig. 2 and derive the results that the coordinates (r_m, θ_m) and (r_n, θ_n) of the same point P are related to each other by

$$r_m = \sqrt{r_n^2 + L_{mn}^2 + 2L_{mn}r_n\cos\theta_n}, \quad \theta_m = \mathcal{F}(r_n, \theta_n, L_{mn}), \tag{2.15}$$

where

$$L_{mn} = a_n - a_m, \tag{2.16}$$

and

$$\mathcal{F}(r,\theta,L) = \arccos\left(\frac{r\cos\theta + L}{\sqrt{r^2 + L^2 + 2Lr\cos\theta}}\right) \tag{2.17}$$

is an analytic function which admits a less geometrically transparent but more convenient form

$$\mathcal{F}(r,\theta,L)=\frac{i}{2}\log\left(\frac{L+re^{i\theta}}{L+re^{-i\theta}}\right). \tag{2.18}$$

Then, analyzing the triangle a_mPa_{m+1} shown in Fig. 2 we see that $P\in I_m$ if and only if $\theta_{m+1}-\theta_m=\pi$, which may be formulated in the coordinates (r_n,θ_n) as

$$P\in I_m\Leftrightarrow\mathcal{F}(r_n,\theta_n,a_n-a_{m+1})-\mathcal{F}(r_n,\theta_n,a_n-a_m)=\pi, \tag{2.19}$$

where $\mathcal{F}(r,\theta,L)$ is the function (2.18).

It is important to note that since $\mathcal{F}(r,\theta,L)$ is an analytic function of r and θ, then Eq. (2.19) determines a two-dimensional multisheet manifold

$$\mathcal{T}_m^{(n)}=\{r_n,\theta_n:\mathcal{F}(r_n,\theta_n,a_n-a_{m+1})-\mathcal{F}(r_n,\theta_n,a_n-a_m)=\pi\}, \tag{2.20}$$

which may be considered as an analytical continuation of the segment I_m described in terms of polar coordinates (r_n,θ_n).

In the coordinates (r_n,θ_n) the line l_n from Eq. (2.12) is described as

$$l_n=\{r_n,\theta_n:r_n>0,\theta_n=\phi\}\quad\text{where }\phi=\pi-\theta_*, \tag{2.21}$$

and the interface conditions (2.13), (2.14) have the form

$$U_n\big|_{l_n^+}-U_n\big|_{l_n^-}=K_ne^{ikr_n},\qquad\left.\frac{\partial U_n}{\partial n}\right|_{l_n^+}=\left.\frac{\partial U_n}{\partial n}\right|_{l_n^-}, \tag{2.22}$$

where

$$K_n=(R_n-R_{n+1})e^{-ika_n\cos\theta_*}. \tag{2.23}$$

Then, seeking $U_n(r,\theta)$ as the product

$$U_n(r,\theta)=u_n(r,\theta)e^{ikr}, \tag{2.24}$$

we arrive at

$$\frac{r^2}{2}\frac{\partial^2u_n}{\partial r^2}+r\left(\frac{1}{2}+ikr\right)\frac{\partial u_n}{\partial r}+\frac{1}{2}\frac{\partial^2u_n}{\partial\theta^2}+\frac{ikr}{2}u_n=0, \tag{2.25}$$

accompanied by the conditions at infinity

$$u_n(r,\theta)=o(1)\quad\text{as } r\to\infty\quad\theta\neq\phi, \tag{2.26}$$

the interface conditions

$$u_n(r,\phi+0)-u_n(r,\phi-0)=K_n,$$

$$\left.\frac{\partial u_n}{\partial\theta}\right|_{\theta=\phi+0}=\left.\frac{\partial u_n}{\partial\theta}\right|_{\theta=\phi-0}, \tag{2.27}$$

and by the boundary conditions

$$\frac{\partial u_n(r,\theta)}{\partial\theta}+ikr\mathcal{B}_n(r,\theta)u_n(r,\theta)=0\quad\text{if }\theta=0,$$

$$-\frac{\partial u_n(r,\theta)}{\partial\theta}+ikr\mathcal{B}_n(r,\theta)u_n(r,\theta)=0\quad\text{if }\theta=\pi, \tag{2.28}$$

where

$$\mathcal{B}_n(r,\theta)=\begin{cases}B_m & \text{if }(r,\theta)\in\mathcal{T}_m^{(n)},\\ 0 & \text{otherwise.}\end{cases} \tag{2.29}$$

Since $\mathcal{T}_m^{(n)}$ are defined by Eq. (2.20) as analytic continuations of the intervals I_m, it is natural to view $\mathcal{B}_n(r,\theta)$ as an analytic continuation of the piecewise constant impedance $B(x)$ from the one-dimensional domain $I=\cup_mI_m$ to a two-dimensional multisheet analytic manifold $\mathcal{T}=\cup_m\mathcal{T}_m^{(n)}$ described in polar coordinates (r_n,θ_n) by Eq. (2.20).

In Ref. 3 we introduced a method for explicit solution of Eq. (2.25) in the finite angle $\alpha<\theta<\beta$ with interface boundary conditions of the type (2.27), and it is easy to see that this method can be straightforwardly applied to the problem (3.20)–(3.24) formulated earlier. As a result, we arrive at the representation of $u_n(r,\theta)$ in the form

$$u_n(r,\theta)=K_n\mathbf{E}\left\{\sum_{\nu=1}^{\infty}\delta_\nu(\phi)\exp\left(\int_0^{\tau_\nu}ik\xi_t\left[\frac{dt}{2}+\mathcal{B}_n(\xi_t,\eta_t)d\lambda_t\right]\right)\right\}, \tag{2.30}$$

the exact meaning of which is explained in the following.

The mathematical expectation $\mathbf{E}$ is computed over the trajectories of the independent random motions ξ_t and η_t referred to hereafter as the radial and the angular motions, respectively.

The radial motion is launched at the time $t=0$ from the position $\xi_0=r$ and is controlled by the stochastic differential equation

$$\xi_0=r,\quad d\xi_t=\xi_tdw_t^1+\xi_t\left(\frac{1}{2}+ik\xi_t\right)dt, \tag{2.31}$$

where w_t^1 is the standard one-dimensional Brownian motion. As shown in Ref. 2 the radial motion ξ_t drifts toward the unreachable point $\xi=i/2k$ and, at any time $t>0$ it is located inside the quadrant $\mathrm{Re}(\xi_t)>0$, $\mathrm{Im}(\xi_t)\geqslant 0$.

The angular motion η_t and its local time λ_t are stochastic processes determined by

$$\eta_0=\theta,\quad d\eta_t=\begin{cases}dw_t^2 & \text{if } 0<\eta_t<\pi\\ -dt & \text{if } \eta_t=0\\ dt & \text{if } \eta_t=\pi,\end{cases} \tag{2.32}$$

$$\lambda_0=0,\quad d\lambda_t=\begin{cases}0 & \text{if } 0<\eta_t<\pi\\ dt & \text{if } \eta_t=0 \text{ or } \eta_t=\pi,\end{cases} \tag{2.33}$$

which show that while η_t is located inside the interval $0\leqslant\eta_t\leqslant\pi$ it runs as a Brownian motion and λ_t remains constant, but when η_t reaches the interval's border the local time λ_t increases and η_t is deterministically reflected back to the segment a distance equal to the increase of λ_t.

The angular motion running inside the interval $[0,\pi]$ touches the fixed point $\eta=\phi$ from Eq. (3.13) at the times $t=\tau_\nu$ enumerated by the index $\nu\geqslant 1$ which determines the factors $\delta_\nu(\phi)$ by the following rules:

$$\delta_\nu(\phi)=\begin{cases}1 & \text{if } \phi<\eta_{\tau_\nu-0} \quad \text{and} \quad \eta_{\tau_\nu+0}<\phi \\ -1 & \text{if } \phi>\eta_{\tau_\nu-0} \quad \text{and} \quad \eta_{\tau_\nu+0}>\phi \\ 0 & \text{otherwise.}\end{cases} \tag{2.34}$$

It is clear that $\delta_\nu=1$ if at the time τ_ν the interface $\eta=\phi_1$ or $\eta=\phi_2$ is crossed from left to right. Similarly, the value $\delta_\nu=-1$ corresponds to the crossing from right to left, and $\delta_\nu=0$ is the case when the interface is touched but not intersected.

Finally, it is important to observe that since the radial motion ξ_t is confined to the first quarter of the complex plane, the absolute value of the exponent in Eq. (2.30) decreases as $\nu\to\infty$ at the rate $O(e^{-\epsilon\nu})$, with $\epsilon>0$, which guarantees the convergence of Eqs. (2.30) and (2.24).

C. Discussion of the probabilistic solution

A rigorous discussion of stochastic differential equations, stochastic integrals, and of local times can be found in the literature on stochastic processes,[5,6] but for our purposes it suffices to view the random motions ξ_t, η_t and the integrals from Eq. (2.30) as the limits as $\Delta t\to 0$ of discrete processes as described in the following.

The radial motion ξ_t that is controlled by the stochastic equation (2.31) can be considered as a sequence of random jumps

$$\xi_t\to\xi_{t+\Delta t}=\xi_t\pm\xi_t\sqrt{\Delta t}+\xi_t\left(\tfrac{1}{2}+ik\xi_t\right)\Delta t, \tag{2.35}$$

following each other with an infinitesimally small time increment $\Delta t\to 0$. Similarly, the angular motion η_t may be approximated by discrete jumps determined by the rule

$$\eta_t\to\eta_{t+\Delta t}=\begin{cases}\eta_t\pm\sqrt{\Delta t} & \text{if } \eta_t\neq 0,\alpha \\ \eta_t-\Delta t & \text{if } \eta_t=0 \\ \eta_t+\Delta t & \text{if } \eta_t=\pi,\end{cases} \tag{2.36}$$

depending on the current position of the moving point. These discrete approximations of the radial and angular random motions are closely related to the possibility of approximating the integrals from Eq. (2.30) by the Riemann sums

$$\int_0^t\xi_s ds\approx\Delta t\sum_{\nu=0}^{\nu\Delta t\leqslant t}\xi_{\nu\Delta t}, \quad \int_0^t\xi_s d\lambda_s^{1,2}\approx\Delta t\sum_{\nu=0}^{\nu\Delta t\leqslant t}\Omega_\nu^{1,2}\xi_{\nu\Delta t}, \tag{2.37}$$

where the factors

$$\Omega_\nu^1=\begin{cases}1 & \text{if } \nu\Delta t=0 \\ 0 & \text{otherwise,}\end{cases} \quad \Omega_\nu^2=\begin{cases}1 & \text{if } \nu\Delta t=\alpha \\ 0 & \text{otherwise,}\end{cases} \tag{2.38}$$

indicate the times when the angular motion η_t is reflected by the boundaries $\eta=0$ and $\eta=\pi$, respectively. Correspondingly, the local times λ_t^1 and λ_t^2 can be approximated as

$$\lambda_t^n=\int_0^t d\lambda_t^n\approx\Delta t\sum_{\nu=0}^{\nu\Delta t\leqslant t}\Omega_\nu^n, \quad n=1,2. \tag{2.39}$$

It is important to emphasize that the discrete motions (2.35) and (2.36) can be regarded as the Euler approximations to the solutions of the stochastic differential equations (2.31) and (2.32), but these approximations should by no means be considered as the sole methods of numerical simulation of the random motions ξ_t and η_t defined by Eqs. (2.31) and (2.32). Similarly, expression (2.37) should not be regarded as the sole method of computing the integrals from Eq. (2.30) which they indeed approximate. The situation here is reminiscent of the classical theory of integration, where the Riemann sums provide a handy interpretation of the integrals while the actual integration is more efficiently done either analytically or by sophisticated and efficient numerical schemes.

Finally, for better understanding of the obtained probabilistic solution it is instructive to analyze how the trajectory of the random motion (ξ_t,η_t) hits a boundary $\theta=0$ or $\theta=\pi$.

For definiteness and simplicity we assume that the index n is fixed which means that we consider only one particular field U_n. Then, the motion (ξ_t,η_t) starts from the real-valued position $(\xi_0,\eta_0)=(r,\theta)$, where (r,θ) are the coordinates of the observer P in the nth coordinate system. Since the angular motion η_t stays on the real interval $(0,\pi)$ and the radial motion ξ_t runs across the complex plane, the trajectory of (ξ_t,η_t) belongs to the space $\mathbb{C}\times\mathbb{R}$ and completely determines the path of the control functions

$$\rho_m(t)=\mathcal{F}(\xi_t,\eta_t,a_n-a_{m+1})-\mathcal{F}(\xi_t,\eta_t,a_n-a_m), \tag{2.40}$$

$$0\leqslant m\leqslant N,$$

which start from the real-valued points $\rho_m(0)=\theta_m-\theta_{m+1}$, but have thereafter complex values all the time except for the instants $t=t_0$, when $\eta_{t_0}=0$ or $\eta_{t_0}=\pi$. At these exceptional times, functions $\rho_m(t)$ take the real values

$$\rho_m(t_0)=\arccos(\pm 1)=\pi l_m, \quad 0\leqslant m\leqslant N, \tag{2.41}$$

where all of the integers l_m vanish, but one, let it be l_{m_0}, equals unity, so that

$$\rho_m(t_0)=\begin{cases}\pi & \text{if } m=m_0 \\ 0 & \text{otherwise,}\end{cases} \tag{2.42}$$

and, therefore, function $\mathcal{B}_n(\xi_{t_0},\eta_{t_0})$ from Eq. (2.29) has the value

$$\mathcal{B}_n(\xi_{t_0},\eta_{t_0})=B_{m_0}. \tag{2.43}$$

After that, the angular motion η_t is reflected back to the interval $(0,\pi)$, and the process continues starting afresh from the position $(\xi_{t_0+dt},\eta_{t_0+dt})$.

D. Example

To illustrate the feasibility of the obtained probabilistic solution (2.30)–(2.34) for practical calculations we conducted numerical simulations of the wave fields in a half-plane $y>0$ with the boundary $y=0$, $-\infty<x<\infty$ which consists of four segments $I_1,\ldots,I_4$ which have constant impedances $B_1,\ldots,B_4$, and which are separated by the three points $a_1=-5$, $a_2=-2$, and $a_3=4$, each of which may be considered as a source of a separate diffracted wave. We considered the three distinct combinations of the impedances shown as follows:

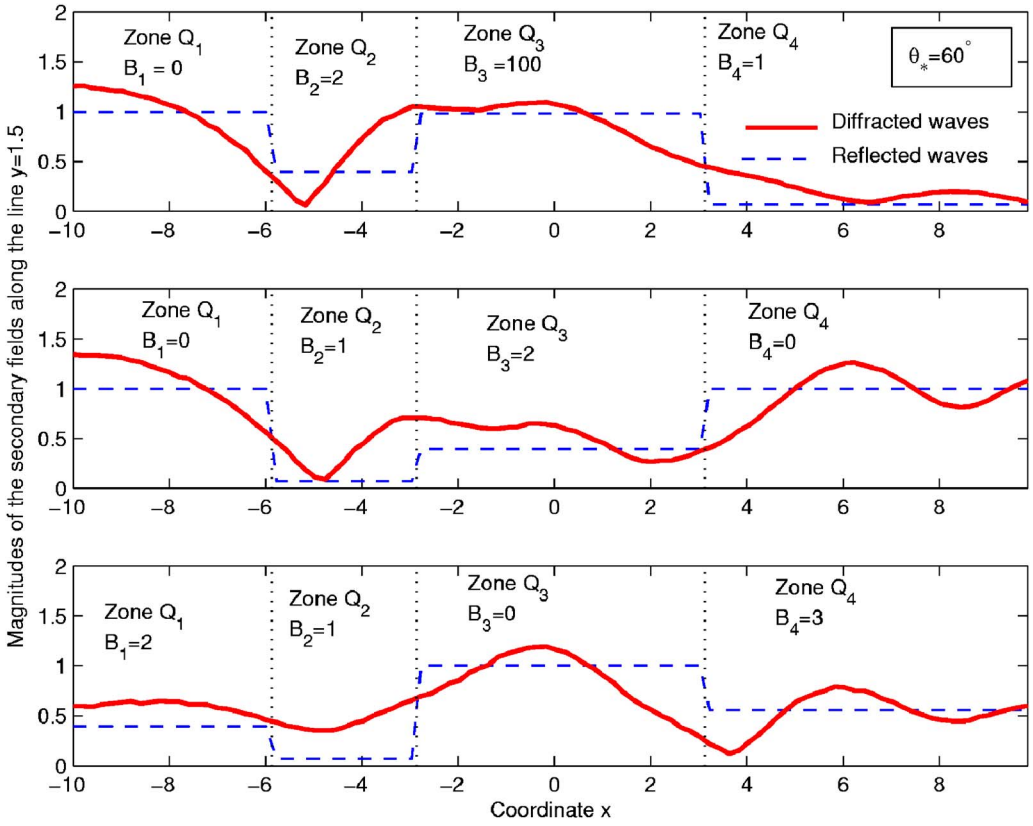


FIG. 3. (Color online) Magnitudes of the secondary wave fields.

	B_1	B_2	B_3	B_4
	$-\infty<x<-5$	$-5<x<-2$	$-2<x<4$	$4<x<\infty$
Case 1	0	2	100	1
Case 2	0	1	2	0
Case 3	2	1	0	3

It should be noted that a vanishing impedance $B=0$ corresponds to the Neumann boundary condition, and a large impedance $B=100$ may be regarded as an approximation to the Dirichlet boundary condition corresponding to the impedance $B=\infty$.

Figures 3 and 4 show the magnitudes and the phases of the diffracted and reflected wave fields generated in the half plane by the plane wave U_* from Eq. (2.1) with the incidence angle $\theta_*=60°$. The dashed lines correspond to the discontinuous reflected wave U_r from Eqs. (2.8) to (2.12) and the solid lines correspond to the continuous secondary wave field $(U_r+U_1+\ldots U_3)$ which consists of the reflected wave and of the three diffracted waves centered at the boundary points $a_1,\ldots,a_3$ where the impedance jumps. Similarly, in Fig. 5 the bold lines show the magnitudes of the continuous total wave fields $U=U_*+U_r+U_1+\cdots+U_3$, and the dashed lines show the magnitudes of the discontinuous geometric fields $U_g=U_*+U_r$, which consist of the incident and reflected waves.

It is instructive to observe that all of the graphs demonstrate the expected periodicity behavior of the wave fields. Indeed, the geometrical optics part predicts that the amplitudes of the reflected field along the line $y=$const depend on the coordinate x as piecewise constant functions, and that the phases of these fields are piecewise linear functions. These features are clearly displayed by the shape of the dashed lines in the above figures. However, due to diffraction, the jumps of the wave fields are smoothed when the diffracted waves are added, and in the case that a geometrical zone is narrow and the reflected field in it is not strong, then the total field in that zone may be overwhelmed by the influence from the neighboring zones. This physically expected prediction is completely confirmed by the presented graphs where the fields in the zone Q_2 are highly influenced by the fields in the zones Q_1 and Q_3.

All of the numerical results presented here were obtained by the straightforward approximation of the mathematical expectation (2.30) by averaging the sample values of the functional computed along the trajectories of 1000 discrete random walks with the time increment $\Delta t=0.01$.

III. DIFFRACTION BY A SEGMENT WITH IMPEDANCE BOUNDARY CONDITIONS

A. Formulation of the problem

Let (x,y) be Cartesian coordinates in a plane with a screen Γ occupying a straight segment of length $L>0$ connecting points $O_1=(a,0)$ and $O_2=(a+L,0)$, as shown in Fig. 6. Then, the problem of diffraction of a plane wave

$$U_* = e^{-ik(x\cos\theta_*+y\sin\theta_*)}, \quad 0\leqslant\theta_*\leqslant\pi, \tag{3.1}$$

by the screen Γ consists of the computation of the bounded solution of the Helmholtz equation

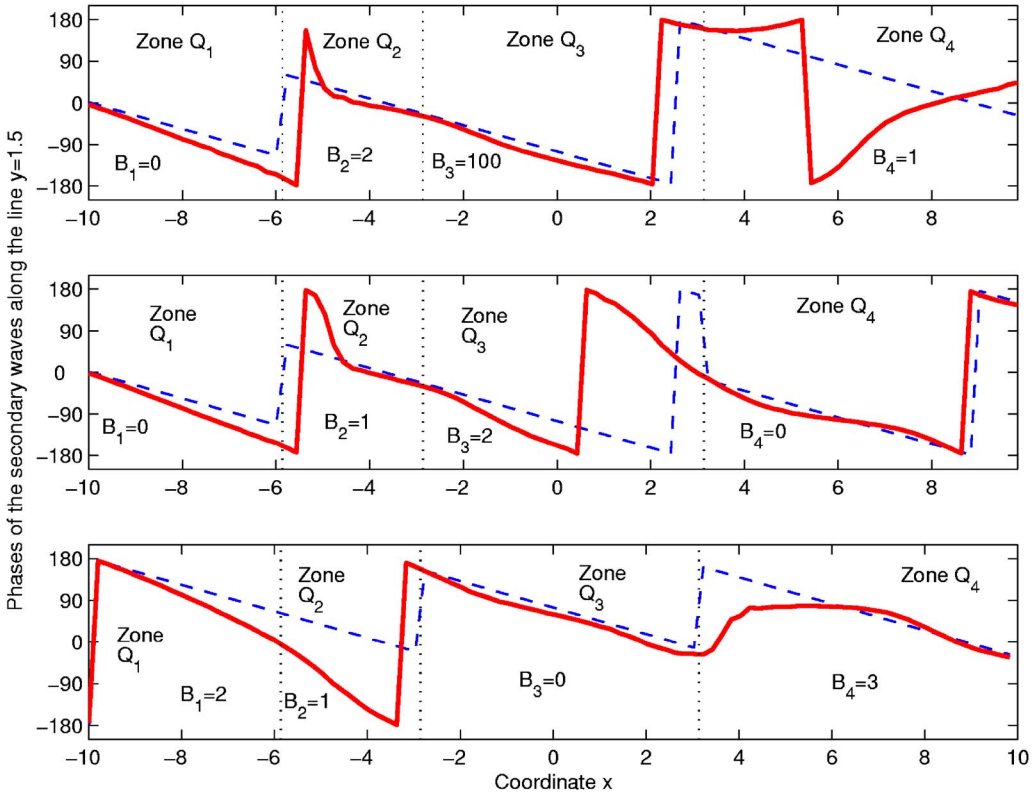


FIG. 4. (Color online) Phases of the secondary wave fields.

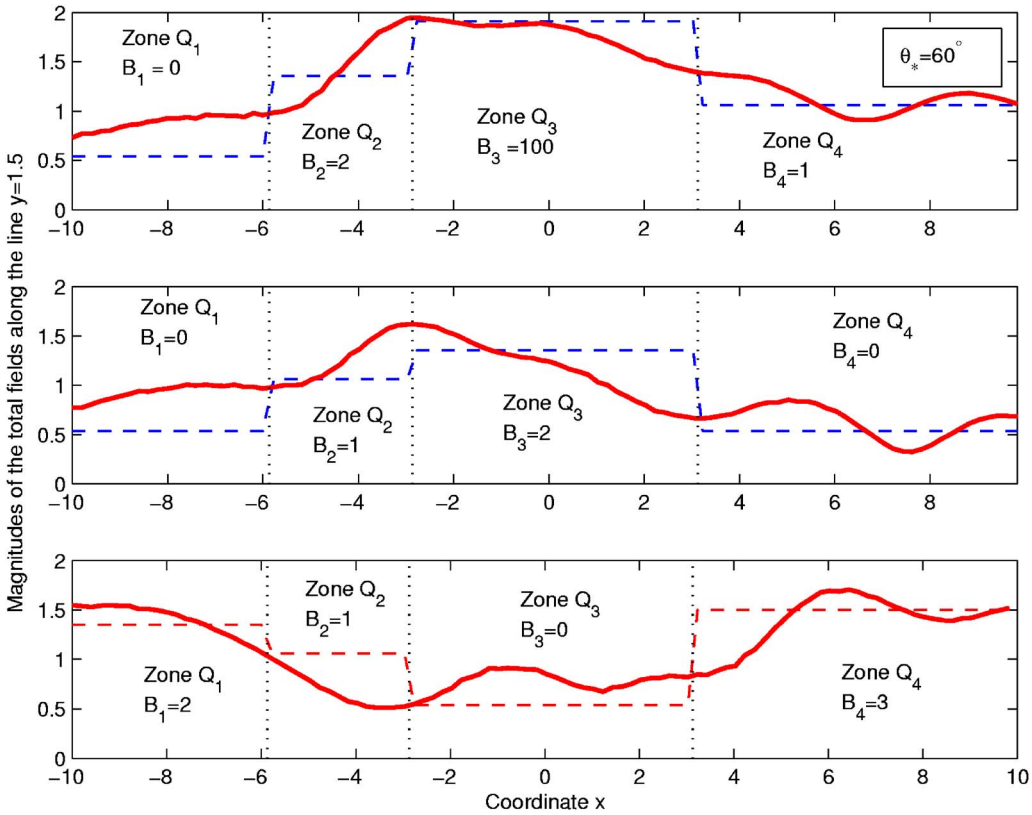


FIG. 5. (Color online) Magnitudes of the total wave fields.

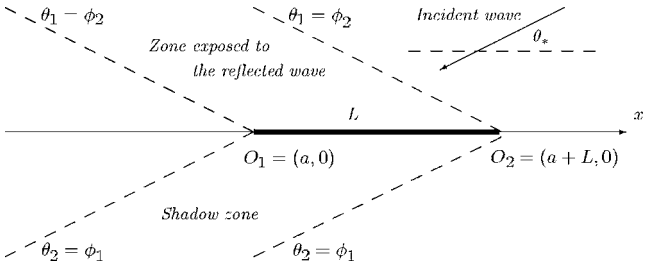


FIG. 6. Diffraction by a segment: geometry of the problem.

$$\nabla^2 U + k^2 U = 0, \tag{3.2}$$

which does not contain waves arriving from infinity except for the incident wave U_* and satisfies the impedance boundary conditions

$$\pm \frac{\partial U}{\partial y} + ikB_{\pm} U \bigg|_{\Gamma_{\pm}} = 0, \tag{3.3}$$

where Γ_+ and Γ_- denote the upper and the lower sides of the screen Γ, which have constant impedances $B_{\pm}$ restricted by

$$\mathrm{Re}(B_{\pm}) \geqslant 0. \tag{3.4}$$

Elementary geometric-optical analysis suggests that the solution of this diffraction problem can be represented as a superposition

$$U = U_i + U_r + U_1 + U_2, \tag{3.5}$$

of the discontinuous incident and reflected waves U_i and U_r complimented by the (also discontinuous) diffracted fields U_1 and U_2.

To describe the components of Eq. (3.5) in detail we introduce the two polar coordinates (r_1, θ_1) and (r_2, θ_2) centered at the vertices O_1 and O_2, shown in Fig. 7. For definiteness we assume that the polar angles increase in the counterclockwise direction from the semi-axis $y=0$, $x>0$.

Straightforward application of formulas (2.15)–(2.18) reveals that the new coordinates are related to each other by

$$r_2 = \sqrt{r_1^2 + L^2 - 2Lr_1 \cos\theta_1}, \quad \theta_2 = \mathcal{F}(r_1, \theta_1, -L), \tag{3.6}$$

$$r_1 = \sqrt{r_2^2 + L^2 + 2Lr_2 \cos\theta_2}, \quad \theta_1 = F(r_2, \theta_2, L), \tag{3.7}$$

where $\mathcal{F}(r, \theta, L)$ is a function from Eq. (2.18). In these coordinates the segment $\Gamma = (O_1, O_2)$ can be described by either of the equivalent equations

$$\Gamma_{\pm}: \quad \theta_1 - F(r_1, \theta_1, -L) = \mp \pi, \tag{3.8}$$

or

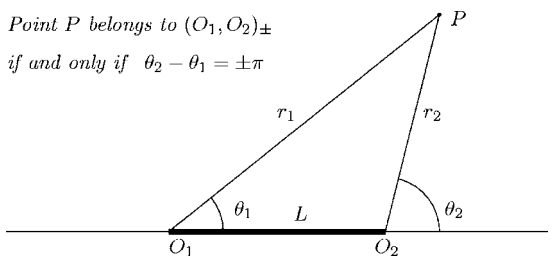


FIG. 7. Polar coordinates (r_1, θ_1) and (r_2, θ_2).

$$\Gamma_{\pm}: \quad \theta_2 - F(r_2, \theta_2, L) = \pm \pi, \tag{3.9}$$

where the upper signs in the right-hand sides correspond to the upper side of Γ, and the lower signs correspond to the lower side of Γ.

Using both of the introduced polar coordinate systems we conveniently describe the incident and reflected components of Eq. (3.5) by the formulas

$$U_i(r_n, \theta_n) = \begin{cases} T_n e^{-ikr_n \cos(\theta_n - \theta_*)} & \text{if } \phi_2 \notin (\theta_1, \theta_2) \\ 0 \quad \text{otherwise}, \end{cases} \tag{3.10}$$

and

$$U_r(r_n, \theta_n) = \begin{cases} T_n K e^{-ikr_n \cos(\theta_n + \theta_*)} & \text{if } \phi_1 \in (\theta_1, \theta_2) \\ 0 \quad \text{otherwise}, \end{cases} \tag{3.11}$$

where

$$T_1 = e^{ika\cos\theta_*}, \quad T_2 = e^{ik(a+L)\cos\theta_*}, \quad K = \frac{\sin\theta_* + B_+}{\sin\theta_* - B_+}, \tag{3.12}$$

and

$$\phi_1 = \pi - \theta_*, \quad \phi_2 = \pi + \theta_* \tag{3.13}$$

are the angles determining the shadow boundaries.

The sum $U_g = U_i + U_r$ of the incident and reflected fields can be viewed as the first approximation to the exact solution which obeys all conditions of the problem except that it is discontinuous along the rays $\theta = \phi_1$ and $\theta = \phi_2$. To remove these discontinuities we require that the diffracted field U_1 obeys the interface conditions

$$\begin{aligned} &U_1(r_1, \phi_1 + 0) - U_1(r_1, \phi_1 - 0) \\ &\quad = T_1 K e^{ikr_1}, \quad \frac{\partial U_1}{\partial \theta_1}\bigg|_{\theta_1 = \phi_1 + 0} = \frac{\partial U_1}{\partial \theta_1}\bigg|_{\theta_1 = \phi_1 - 0}, \\ &U_1(r_1, \phi_2 + 0) - U_1(r_1, \phi_2 - 0) \\ &\quad = T_1 e^{ikr_1}, \quad \frac{\partial U_1}{\partial \theta_1}\bigg|_{\theta_1 = \phi_2 + 0} = \frac{\partial U_1}{\partial \theta_1}\bigg|_{\theta_1 = \phi_2 - 0}, \end{aligned} \tag{3.14}$$

formulated in the coordinates (r_1, θ_1), and the field U_2 obeys similar conditions,

$$\begin{aligned} &U_2(r_2, \phi_1 + 0) - U_2(r_2, \phi_1 - 0) \\ &\quad = -T_2 K e^{ikr_2}, \quad \frac{\partial U_2}{\partial \theta_2}\bigg|_{\theta_2 = \phi_1 + 0} = \frac{\partial U_2}{\partial \theta_2}\bigg|_{\theta_2 = \phi_1 - 0}, \\ &U_2(r_2, \phi_2 + 0) - U_2(r_2, \phi_2 - 0) \\ &\quad = -T_2 e^{ikr_2}, \quad \frac{\partial U_2}{\partial \theta_2}\bigg|_{\theta_2 = \phi_2 + 0} = \frac{\partial U_2}{\partial \theta_2}\bigg|_{\theta_2 = \phi_2 - 0}, \end{aligned} \tag{3.15}$$

formulated in the coordinates (r_2, θ_2). We also require that the diffracted fields U_1 and U_2 satisfy the Helmholtz equations

$$\nabla^2 U_n(r_n,\theta_n)+k^2U_n(r_n,\theta_n)=0,\quad \theta\neq\pi\pm\theta_*\quad (n=1,2), \tag{3.16}$$

accompanied by the radiation conditions

$$e^{-ikr}U_n(r_n,\theta_n)=o(1)\quad \text{as } r\to\infty,\quad \theta_n\neq\pi\pm\theta_*, \tag{3.17}$$

and by the boundary conditions

$$\pm\frac{\partial U_1}{\partial\theta_n}+ikB_\pm U_1=0\quad \text{as } \mathcal{F}(r_1,\theta_1,-L)-\theta-1=\pm\pi,$$
$$\mp\frac{\partial U_2}{\partial\theta_n}+ikB_\pm U_2=0\quad \text{as } \theta_2-\mathcal{F}(r_2,\theta_2,L)=\pm\pi. \tag{3.18}$$

Finally, for $r_n>L$ the fields $U_n(r_n,\theta_n)$ must be 2π-periodic with respect to the angle θ_n.

B. Computation of the diffracted fields

We seek the 2π-periodic fields $U_1(r,\theta)$ and $U_2(r,\theta)$ in the form

$$U_n(r,\theta)=T_ne^{ikr}\sum_{m=-\infty}^{\infty}u_n(r,\theta+2\pi m),\quad n=1,2, \tag{3.19}$$

where the amplitudes $u_n(r,\theta)$, with $n=1,2$, are bounded solutions of

$$\frac{r^2}{2}\frac{\partial^2u_n}{\partial r^2}+r\left(\frac{1}{2}+ikr\right)\frac{\partial u_n}{\partial r}+\frac{1}{2}\frac{\partial^2u_n}{\partial\theta^2}+\frac{ikr}{2}u_n=0, \tag{3.20}$$

considered in the domain $r>0$, $-\infty<\theta<\infty$ bounded by the lines

$$\mathfrak{G}_\pm^1:\quad F(r,\theta,-L)-\theta=\pm\pi,$$
$$\mathfrak{G}_\pm^2:\quad \theta-F(r,\theta,+L)=\pm\pi, \tag{3.21}$$

which may be considered as extensions of the segments $\Gamma_\pm$ to the domain with infinite angular range. Moreover, since the function $\mathcal{F}(r,\theta,L)$ is analytic, domains $\mathfrak{G}_\pm^n$ can be considered as two-dimensional multisheet manifolds which analytically continue the segments $\Gamma_\pm$.

To guarantee that the series (3.19) satisfies the boundary conditions (3.18) we require that the amplitudes $u_n(r,\theta)$ satisfy the boundary conditions

$$\pm\frac{\partial u_n(r,\theta)}{\partial\theta}+ik\mathcal{B}_n(r,\theta)u_n(r,\theta)=0\quad \text{when } (r,\theta)\in\mathfrak{G}_\pm^n, \tag{3.22}$$

where, similar to Eq. (2.29), functions

$$\mathcal{B}_n(r,\theta)=\begin{cases}B_+ & \text{if } (r,\theta)\in\mathfrak{G}_+^n,\\ B_- & \text{if } (r,\theta)\in\mathfrak{G}_-^n,\end{cases} \tag{3.23}$$

may be treated as analytical continuations of the piecewise constant impedance from the real one-dimensional boundary $\Gamma=\Gamma_+\cup\Gamma_-$ to the two-dimensional multivalued analytical surface $\mathfrak{G}_+^n\cup\mathfrak{G}_-^n$ from Eq. (3.21).

We also replace the radiation conditions (3.17) by the conditions

$$u_n(r,\theta)=o(1)\quad \text{as } r\to\infty,\quad \theta\neq\phi_{1,2}, \tag{3.24}$$

where ϕ_1, ϕ_2 are the angles from Eq. (3.13). Finally, the interface conditions (3.14) and (3.15) are replaced by the conditions

$$u_n(r,\phi_1+0)-u_n(r,\phi_1-0)=K,$$
$$\left.\frac{\partial u_n}{\partial\theta_n}\right|_{\theta_n=\phi_1+0}=\left.\frac{\partial u_n}{\partial\theta_n}\right|_{\theta_n=\phi_1-0},$$
$$u_n(r,\phi_2+0)-u_n(r,\phi_2-0)=1,$$
$$\left.\frac{\partial u_n}{\partial\theta_n}\right|_{\theta_n=\phi_2+0}=\left.\frac{\partial u_n}{\partial\theta_n}\right|_{\theta_n=\phi_2-0}, \tag{3.25}$$

where $\phi_{1,2}$ have the same meaning as in Eq. (3.25).

The obtained problem (3.20)–(3.24) is essentially identical to the already solved problem (2.25)–(2.29), and using this analogy we readily arrive at

$$u_n(r,\theta)=\mathbf{E}\left\{\sum_{\nu=1}^{\infty}[K\delta_\nu(\phi_1)+\delta_\nu(\phi_2)]\times\exp\left(\int_0^{\tau_\nu}ik\xi_t\left[\frac{dt}{2}+\mathcal{B}(\xi_t,\eta_t)d\lambda_t\right]\right)\right\}, \tag{3.26}$$

which retains most of the notation from Eqs. (2.30)–(2.34). More precisely, the radial random motion ξ_t is controlled by Eq. (2.31). In the case $n=1$, the angular motion η_t is controlled by

$$\eta_0=\theta,\quad d\eta_t=\begin{cases}-dt & \text{if } \mathcal{F}(\xi_t,\eta_t,-L)-\eta_t=\pi\\ dt & \text{if } \mathcal{F}(\xi_t,\eta_t,-L)-\eta_t=-\pi\\ dw_t^2 & \text{otherwise,}\end{cases} \tag{3.27}$$

and in the case $n=2$ the motion η_t is defined by

$$\eta_0=\theta,\quad d\eta_t=\begin{cases}-dt & \text{if } \eta_t-\mathcal{F}(\xi_t,\eta_t,L)=\pi\\ dt & \text{if } \eta_t-\mathcal{F}(\xi_t,\eta_t,L)=-\pi\\ dw_t^2 & \text{otherwise,}\end{cases} \tag{3.28}$$

which show that most of the time the motion η_t runs along the real line $-\infty<\theta<\infty$ as the standard Brownian motion w_t^2, but when the point (ξ_t,η_t) reaches one of the manifolds $\mathfrak{G}_\pm^n$ from Eq. (3.21), the motion η_t is deterministically reflected. Finally, the index ν in Eq. (3.26) enumerates the times t_ν when the angular motion η_t controlled by Eq. (2.32) touches either of the two rays $\theta=\phi_1$ or $\theta=\phi_2$ from Eq. (3.13).

C. Example

To illustrate the obtained probabilistic solution of the problem of diffraction by a finite screen, we conducted numerical simulations of the wave fields generated by the incident plane wave $U_*=e^{-ir\cos(\theta-\theta_*)}$ by a screen Γ occupying the segment $\Gamma=\{x,y:|x|<L/2,y=0\}$ with different impedances on its sides. We considered three different configurations

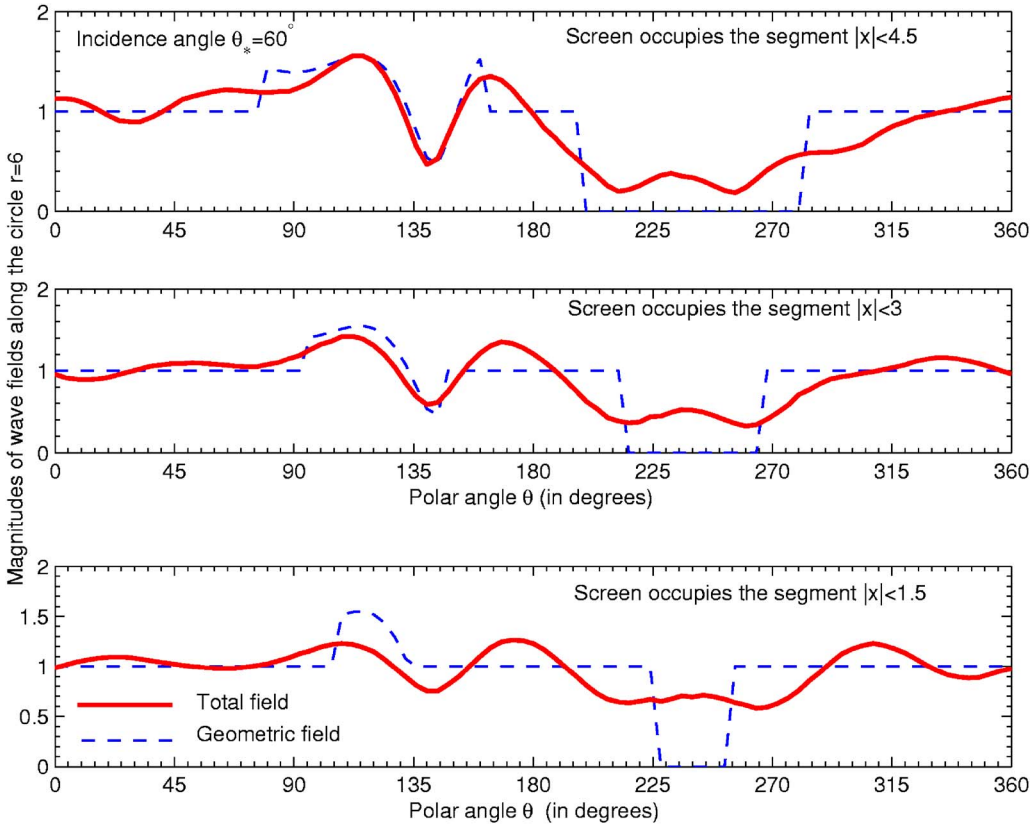


FIG. 8. (Color online) Magnitudes of the total and geometric wave fields.

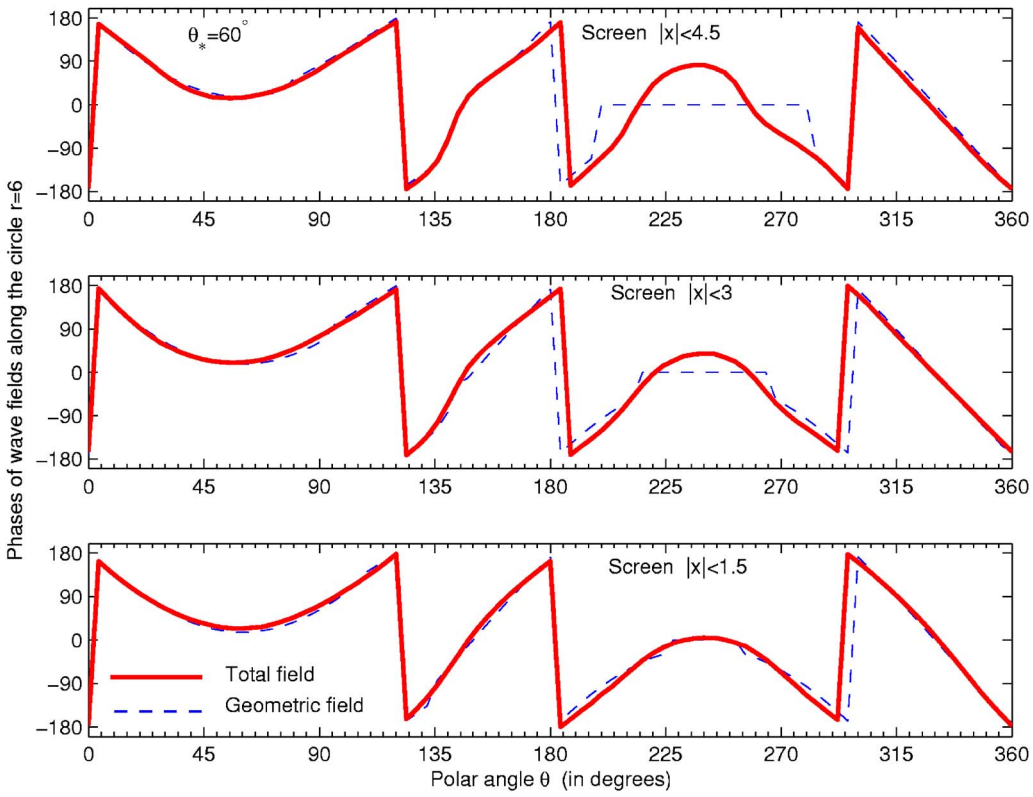


FIG. 9. (Color online) Phases of the total and geometric wave fields.

TABLE I. Decay of $u_{1,2}(r,\theta+2\pi m)$ in the case $L=1.5$.

	$m=0$	$m=-1$	$m=1$	$m=-2$	$m=2$
$\lvert u_1(r,\theta+2\pi m)\rvert$	0.4897	0.0227	0.0066	0.0010	0.0005
$\lvert u_2(r,\theta+2\pi m)\rvert$	0.5072	0.0746	0.0187	0.0035	0.0013

with the segments of the length $L=3$, $L=5$, and $L=9$, which are close to half the wavelength, one wavelength, and 1.5 times the wavelength (2π). In all cases, the incident angle was set to $\theta_*=60°$ and the impedances of the screen were set as

$$B_+=2, \quad B_-=1, \tag{3.29}$$

where the sign "±" correspond to the upper and the lower sides of the screen.

Figures 8 and 9 show the magnitudes and the phases of the total and geometric fields along the circle $\sqrt{x^2+y^2}=6$ centered at the middle of the screen. The dashed lines correspond to the discontinuous geometric field $U_g=U_i+U_r$ described by Eq. (3.5), and the solid lines correspond to the continuous total fields. Additionally, Table I shows the maximal values of the magnitudes of the first few terms of the series (3.19) in the case when $L=1.5$. These terms admit the physical interpretation as waves making several revolutions around the screen, and the presented data clearly demonstrate the convergence of the series.

It is instructive to observe that all of the graphs demonstrate expected periodicity of the wave fields. Moreover, they clearly demonstrate the expected behavior of the total wave fields in the presence of the screens. Indeed, in the absence of any screen, the total field along the circle $\sqrt{x^2+y^2}=6$ would have a constant magnitude $|u_*|=1$ and a harmonic phase given (in radians) by the formula $\arg(u_*)=\pi\cos(\theta-\theta_*)$. In the presence of a screen one expects the total wave field to have a low magnitude in the shadow area behind the screen, and that it should have higher magnitude above the screen in the area illuminated by the reflected waves. These predictions are confirmed by the dashed lines which demonstrate the darkness in the shadow zone around the ray $\theta\approx 230°$ and the typical interference in the sector around the ray $\theta\approx 140°$. However, due to diffraction, the jumps of the wave fields should not be sharp as predicted by the geometrical optics but should be smoothed by the diffracted waves. This physically justified prediction is completely confirmed by the smoothness of the solid lines which represent total fields and are computed by the developed probabilistic method. Additionally, tracing the differences between the bottom diagrams and their upper counterparts, one may observe in the case of the shorter screen (bottom diagrams) that the shadow is not so deep and the interference is not so observable as in the cases of longer screens.

All of the numerical results were obtained by the approximation of the mathematical expectation (3.26) by averaging 1000 sample values of the functional computed along the discrete random motions with the time step $\Delta t=0.01$.

IV. CONCLUSION

The version of the random walk method developed in Refs. 2 and 3 made it possible to obtain explicit solutions of problems of diffraction in wedges with faces of constant impedance. Such domains are characterized by the presence of only one diffraction point located at the vertex of the wedge. Here we extend the method to problems of wave propagation with several diffraction points. For transparency, the present paper is restricted to diffraction by collinear structures, but it may be straightforwardly applied to problems of diffraction by general polygonal objects, which will be considered in a separate paper.

As mentioned in Ref. 3, the probabilistic solutions provided by the random walk method are local in the sense that they make it possible to compute functions of interest at individual points without computing them on dense meshes. Such solutions admit natural parallel computations and supply meaningful physical interpretations which complement elementary models of wave propagation.

ACKNOWLEDGMENTS

This research was supported by NSF Grant No. CMS-0408381 and by the DARPA/AFOSR Contract No. FA 9550-0501-0111.

[1] J. M. L. Bernard, Q. J. Mech. Appl. Math. **58**, 383–418 (2005).
[2] B. V. Budaev and D. B. Bogy, J. Acoust. Soc. Am. **114**, 1733 (2003).
[3] B. V. Budaev and D. B. Bogy, IEEE Trans. Antennas Propag. **53**, 2073 (2005).
[4] G. D. Maliuzhinets, Acoust. J. **1**, 144 (1955).
[5] M. Freidlin, *Functional Integration and Partial Differential Equations*, The Annals of Mathematics Studies No. 109 (Princeton University Press, Princeton, NJ, 1985).
[6] K. Ito and H. P. McKean, Jr., *Diffusion Processes and Their Sample Paths* (Springer, Berlin, 1965).

Causal analysis of transient viscoelastic wave propagation

André Moura[a)]
Lab. Voor Akoestiek en Thermische Fysica Department of Natuurkunde en Sterrenkunde, K. U. Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium

(Received 9 May 2005; revised 17 November 2005; accepted 17 November 2005)

Dispersion due to internal frictions as wave propagates is a consequence of the second principle of thermodynamics. When the wavelengths are several times higher than the mesoscopic inhomogeneities, internal diffractions can be ignored and the propagation medium can then be considered as a continuum at the scale of these wavelengths. Here, we consider the dissipation mechanism due to viscosity only. By mean of Laplace transforms both on time and space, a causal analysis leads us to a closed-form solution, which we think is the simplest analytical form. This is illustrated by searching the viscoelastic Green's function associated with the horizontal shear wave generated by a uniform impulsive line source in an infinite homogeneous medium, whose example is almost mathematically equivalent to the study of the scalar wave generated in viscous fluid. The described method is thus restricted to one-wave propagation problems and is probably not generally applicable when the source generates several waves. In the course to obtain a transient analytical expression of pulsed wave through a dispersive medium, this study proposes a method for transient cylindrical waves, while most previous methods concern plane waves. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151769]

PACS number(s): 43.20.Bi, 43.20.Px, 43.35.Mr [TDM] Pages: 751–755

I. INTRODUCTION

Transient wave propagation has given rise to considerable publications. Although many explicit solutions were obtained with the help of the Laplace's transform, most of them concerned nondispersive wave propagation. To account for the dispersion leads to more calculations, as inversion of the Laplace's transform is in these cases more complicated. However, except in structured media at microscopic scalelike crystals, dispersion occurs almost all the time in artificial or natural media. We are concerned here with anelastic waves. In different research areas such as ultrasounds in industrial materials,[1] audible sound waves in air or seismic waves propagating in earth crest,[2] and inherent internal frictions are the cause of dispersion. In many realistic and complex problems such as fast crack growth in materials science[3,4] and fault advance prior to earthquake in seismology,[5,6] the knowledge of canonical analytical solutions is a prerequisite before launching numerical simulation.[7] Solving such boundary initial values problems is a difficult task due to the fast moving boundaries. Indeed, crack tip interacts in time with the emitted waves due to the fast advance with velocities comparable to Rayleigh, shear, or dilatational wave, depending on Modes I-, II- or III-crack propagation.[8,9] These problems are classically solved from finite element-based methods but no one is still fully satisfactory.

As a basic prerequisite, the Green's functions are necessary and take a fundamental importance. In elastodynamics, a type of such functions concerned nondispersive wave generated by a line or point source[10] and propagating in unbounded media. In this paper, we search for such a function but for viscoelastic media.

Only a few works in the past were devoted to the transient solution of viscoelastic wave equations. Most of them were based on the theory of propagating surfaces of discontinuity, the so-called method of characteristics,[11,12] whereas a few ones referred to De Hoop modification of Cagniard's method.[13–15] It is also worth mentioning a method constructed from the Kramer-Kroenig relationships,[16,17] which is however formulated from physical insight and is not fully mathematically grounded. One reason why so many solving methods of resolution exist is the difficulty to invert a double Laplace transform both in time and space. Since there are two independent complex variables, the integration in the complex plane is then a serious problem. This is precisely the object of this paper. As only one Laplace transform is involved, for one-dimensional propagation, analytical solutions are sometime available; see for instance the wave propagation in porous media.[18] Also in the harmonic case, one inversion only is needed, and the use of conformal transformation leads to a possible asymptotic analysis.[19] But the works of Ludvig and Levin[20] on transient propagation in viscous fluid are certainly the most similar studies to the one presented here, which are also based on the theory of integration of the complex variable.

For the sake of simplicity and without lost of generality, we consider a bidimensional problem instead of a tridimensional one.

II. PROBLEM FORMULATION

Let us consider horizontal shear (SH) waves generated by a uniform impulsive line source in an infinite viscoelastic medium. The points of the space are represented in a Cartesian coordinate system $(Oxyz)$. Thus the body force can be written for a line source along the Oy axis

[a)]Electronic mail: Andre.Moura@fys.kuleuven.be

$$\mathbf{f} = [0, F\delta(x)\delta(z)\delta(t), 0], \tag{1}$$

where δ denotes the Dirac delta function, t the time, and the constant F a force per unit length. Hence only the y component of displacement $u(x,z,t)$ is excited and obeys the wave equation[2]

$$\rho\frac{\partial^2 u}{\partial t^2} - \frac{\partial}{\partial t}\left[\mu(t) * \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2}\right)\right] = F\delta(x)\delta(z)\delta(t), \tag{2}$$

where ρ is the mass density, $\mu(t)$ the relaxation modulus associated with the Lamé elastic constant $\mu_0 = \mu(0)$, and the asterisk represents the time convolution.

One shall solve Eq. (2) by means of integral transforms, i.e., one-sided and double-sided Laplace transforms on time and space, respectively, and use Cauchy integral theorem for inversion purpose.

III. DEFINITION OF LAPLACE TRANSFORMS

It should be noticed that the use of two successive Laplace transforms, on space and on time, results in a bidimensional complex set. The extension of Cauchy's theory of complex integration to such a set is still an open research subject in mathematics, and thus applications in physics are still limited. In this work, we shall use a special integral transform, which enables us to consider successively two one-dimensional complex sets.

The one-sided Laplace transform on time is defined as

$$\bar{u}(x,z,s) = \int_0^\infty u(x,z,t)\exp(-st)dt, \tag{3}$$

with $\mathrm{Re}(s) \geqslant 0$ that makes the integral regular for positive times. By analytical continuation, its inverse is given with the Mellin inversion formula by

$$u(x,z,t) = \frac{1}{2i\pi}\int_{\beta_s} \bar{u}(x,z,t)\exp(st)dt, \tag{4}$$

where β_s means a symmetrical path in the complex s plane with respect to the real s axis and goes to infinity as the modulus $|s|$ tends to infinity with all singularities on the left of β_s. As a special case we can choose the infinite straight line $)a-i\infty,\ a+i\infty($, with a real, which is often considered in mathematical proofs because of its simplicity.[21] Note that the large choice of possible paths β_s is made possible as Lerch's theorem ensures the uniqueness of the original under minor constraints, which are always verified in physics and satisfied by making the right choice in mathematics.

Thus, given $\bar{u}$, by insight and with regard to Eq. (2), one defines the double-sided Laplace transform on x coordinate as

$$\tilde{\bar{u}}(k,z,s) = \int_{-\infty}^\infty \bar{u}(x,z,s)\exp(f(s)kx)dx, \tag{5}$$

with its inverse

$$\bar{u}(x,z,s) = \frac{f(s)}{2i\pi}\int_{\beta_k} \tilde{\bar{u}}(k,z,s)\exp(-f(s)kx)dk, \tag{6}$$

where

$$f(s) = \sqrt{\mu_0 s/\bar{\mu}(s)}, \tag{7}$$

and $\bar{\mu}(s)$ represents the Laplace transform of $\mu(t)$. The bivalued square root function is made single valued by choosing the branch cut so that the real part of the radical is positive, while the domain of analyticity, where these integrals are regular, will be subsequently defined.

What we know from the physics of the problem is that the integrals above exist when the integration path lies along the vertical imaginary axis. Such a path constitutes a so-called Bromwich's path which is generally completed with a large semicircle in the right half-plane of the complex plane centered at the origin along where Jordan's lemma is always verified to take account for the causality, and with a more sophisticated one that avoids all singularities toward the left half-plane to calculate the displacement field in the wake of wave front. The whole closed path thus formed is often called Bromwich's contour. If one integrates along the imaginary s axis with the change of variable $s=-i\omega$ then the one-sided Laplace transform is changed into a Fourier transform of parameter ω, which represents nothing but the pulsation that can be associated to a harmonic wave. It is then well-known that the solution of Eq. (2) is written as a Weyl's integral,[22] which expresses a spherical (cylindrical) wave emanating from a point (line) source as a superposition of plane waves.

IV. TRANSFORM OF THE WAVE EQUATION

From the well-known correspondence principle in viscoelasticity coupled with the definition of the integral transform in Eq. (6), simple algebraic manipulations lead to the solution of Eq. (2) in the form of the following Weyl-type integral

$$\bar{u}(x,z,s) = \frac{1}{2i\pi}\frac{F}{2}g(s)\int_{\beta_k} \frac{\exp[-f(s)(\eta(k)|z| + kx)]}{\eta(k)}dk, \tag{8}$$

and

$$u(x,z,t) = \frac{1}{2i\pi}\frac{F}{2}\int_{\beta_s} g(s)\exp(st)ds\frac{1}{2i\pi} \times \int_{\beta_k} \frac{\exp[-f(s)(\eta(k)|z| + kx)]}{\eta(k)}dk, \tag{9}$$

with

$$\eta(k) = \sqrt{1/c^2 - k^2}, \quad g(s) = 1/[s\bar{\mu}(s)], \tag{10}$$

where $c = \sqrt{\mu_0/\rho}$ is the elastic shear wave celerity. It is noteworthy at this step to make some observation about how Eq. (8) can be simplified whether one considers elastic media. In this case, $\bar{\mu}(s) = \mu_0/s$ and the displacement field is then

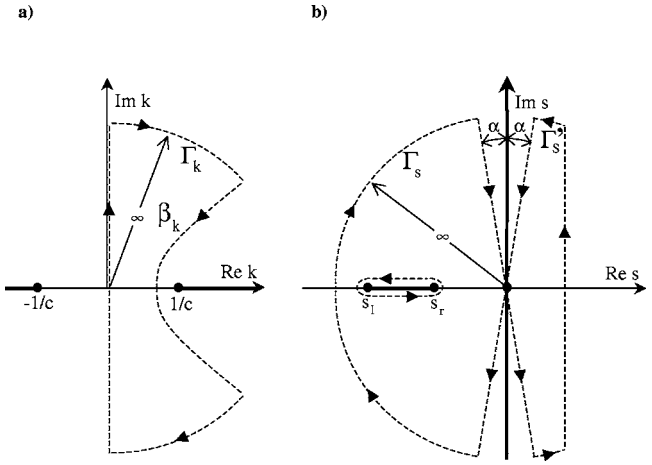


FIG. 1. Bromwich's contours to perform the inverse Laplace transforms. The left- (right)-hand side represents the complex $k(s)$ plane where the contours are in dotted line. The arrows show the integration path. The vertical dotted lines represent original Bromwhich's paths, the paths Γ_k, Γ_s, and Γ_s' represent a large arc of circles centered at the origin where along which Jordan's lemma is applied, while the remaining path β_k is defined by Eq. (12). The two symmetrical straight lines crossing each other on the origin of the complex s plane tend to be vertical as the angle α tends to 0. The lines in bold represent the branch cuts of $\eta(k)$ and $f(s)$, and the black dots their associated branch points.

$$u_e(x,z,t) = \frac{F}{2\mu_0}\frac{1}{2i\pi}\int_{\beta_s}\exp(st)ds\frac{1}{2i\pi} \times \int_{\beta_k}\frac{\exp\{-s[\eta(k)|z|+kx]\}}{\eta(k)}dk. \tag{11}$$

It is well-known in seismology that Cagniard's change of variable[13,23,24]

$$\tau = \eta(k)|z| + kx, \tag{12}$$

which defines here β_k [see Fig. 1(a)] leads to

$$u_e(r,t) = \frac{F}{2\mu_0}\frac{1}{2i\pi}\int_{\beta_s}\exp(s(t-\tau))ds\int_0^\infty\frac{1}{\pi} \times \frac{H(\tau - r/c)}{\sqrt{\tau^2-(r/c)^2}}d\tau, \tag{13}$$

where

$$r = \sqrt{x^2+z^2}, \tag{14}$$

and H denotes the Heaviside function. Note that like β_s, β_k is necessarily symmetrical with respect to the real axis, but since negative values are considered in the space domain then singularities can occur on both sides of β_k.

Now, the identity

$$\delta(t-\tau) = \frac{1}{2i\pi}\int_{\beta_s}\exp[s(t-\tau)]ds \tag{15}$$

gives rise to the well-known final expression after convolution

$$G_e(r,t) = \frac{1}{2\pi\mu_0}\frac{H(t-r/c)}{\sqrt{t^2-(r/c)^2}}, \tag{16}$$

which is nothing but the elastic Green's function. Note that this method of resolution is closely related to the approach initiated by Strick.[24] We will follow next this approach to simplify Eq. (8) but we need first to define the domain of analyticity of the integrand.

V. PROBLEM SOLUTION

In order to avoid any cumbersome calculations, and for the sake of simplicity, one considers the special case of linear viscoelastic medium. Thus, let us consider the relaxation function,[2]

$$\mu(t) = \frac{\mu_0}{\mu_0+\mu_1}[\mu_1 + \mu_0\exp(-t/\tau_0)], \tag{17}$$

where μ_0, μ_1, and τ_0 are constants, with the first corresponding to the Lamé elastic constant, and the last to the relaxation time. Its Laplace transform is

$$\bar{\mu}(s) = \frac{\mu_0}{\mu_0+\mu_1}\left(\frac{\mu_1}{s}+\frac{\mu_0}{s+1/\tau_0}\right). \tag{18}$$

It follows:

$$f(s) = \sqrt{s^2}\sqrt{\frac{s-s_l}{s-s_r}}, \quad g(s) = \frac{1}{\mu_0}\frac{s-s_l}{s-s_r}, \tag{19}$$

where

$$s_l = \frac{-1}{\tau_0}, \quad s_r = \frac{-\mu_1}{\tau_0(\mu_0+\mu_1)}. \tag{20}$$

From our definition, the radicals of $\eta(k)$ and $f(s)$ have the branch cuts $)-\infty,-1/c]\cup[1/c,\infty($ and $[s_l,s_r]\cup\{)-i\infty,0]\cup[0,i\infty(\}$ in the complex k and s planes, respectively. It results the definition of $\sqrt{s^2}$ such that

$$\sqrt{s^2} = s(-s) \quad \text{if } \mathrm{Re}(s) \geqslant 0 (\leqslant 0). \tag{21}$$

The use, once again, of the change of variable defined in Eq. (12) yields

$$u(r,t) = \frac{1}{2i\pi^2}\frac{F}{2}\int_{r/c}^\infty\frac{H(t-r/c)}{\sqrt{\tau^2-(r/c)^2}}\int_{\beta_s}g(s)\exp(-f(s)\tau+st) \times ds d\tau. \tag{22}$$

The integrand in Eq. (22) is analytic everywhere in the complex s plane except on the branch cuts mentioned above. Thus, given $\tau > t$, one can see that the inner integral is regular on the right of β_s as long as this path crosses the real s axis on the right of the imaginary s axis. In turn, by closing the open path β_s with a large semicircle on the right-hand side centered at the origin, and applying Jordan's lemma, calculations from Cauchy's theorem show that the displacement field is null as $\tau > t$ since the integrand is holomorphic on this side and $f(s) \to s$ as $|s| \to \infty$. Values of $\tau \leqslant t$ are now considered in virtue of the principle of analytical continuation. That makes possible to define the integrand over the whole complex s plane, where now the integrand represents a meromorphic function since a branch cut lies on the left-

hand side of β_s in the negative complex s planes. By closing β_s with a semicircle oriented now toward the left-hand side by passing through the imaginary axis on the origin, [see Fig. 1(b)], and wrapping the horizontal cut, one obtains after algebraic manipulations and by applying Cauchy's theorem to the Bromwich's contour thus formed, the following dimensionless viscoelastic Green's function:

$$g_\nu(r,t) = G_\nu(r,t)/A = BH(t'-1)\int_1^{t'} d\tau' \times \int_0^\infty \frac{\nu(\chi)\exp[-u(\chi)t')\sin(\chi u(\chi)\tau']}{\sqrt{\tau'^2-1}} d\chi, \quad (23)$$

where

$$u(\chi) = \frac{s_l' + s_r'\chi^2}{1+\chi^2}, \quad v(\chi) = \frac{\chi^3}{(1+\chi^2)^2}, \quad A = \frac{F}{2\pi\mu_0 t_r},$$

$$B = \frac{2}{\frac{\tau_0}{t_r}\left(1+\frac{\mu_1}{\mu_0}\right)\pi}, \quad (24)$$

$$s_l' = -s_l = \frac{t_r}{\tau_0}, \quad s_r' = -s_r = \frac{\mu_1}{(\mu_0+\mu_1)}\frac{t_r}{\tau_0}, \quad \tau' = \frac{\tau}{t_r},$$

$$t' = \frac{t}{t_r}, \quad t_r = \frac{r}{c}. \quad (25)$$

This final expression only takes into account the branch integral along the branch cut located on the real s axis. Indeed, it is expedient to see that Jordan's lemma is verified in all cases except for $\tau \leqslant t$ along the path Γ_s'. But this last exception is now considered.

Indeed, this comes from the fact that $g(s) \to 1$ as $|s| \to \infty$ and not toward 0. Hence, it is not possible to conclude on the convergence of the integral along the arc of circle Γ_s'. However, as a matter of fact, the part of the integration contour to the right of the imaginary axis in Fig. 1(b) may be considered as a closed contour without any enclosed singularies, so according to Cauchy's theorem its contribution is zero. Consequently, only the closed part to the left of the imaginary axis matters, and this finally yields the integral around the branch cut.

A particular attention is now given to the branch points s_l and s_r, which however as one shall see do not give to any contribution. This is obvious for the left one since s_l is a zero of the function g, but it is not for the right one s_r since it is a simple pole of g. But as it is next shown, this unessential singularity is removed with the exponential convergence of the integrand. Indeed, if one considers the inner integral of Eq. (22) along the small circle path around s_r

$$s = s_r + \epsilon\exp(i\theta), \quad \theta \in [0, 2\pi], \quad \epsilon \to 0, \quad (26)$$

then one obtains

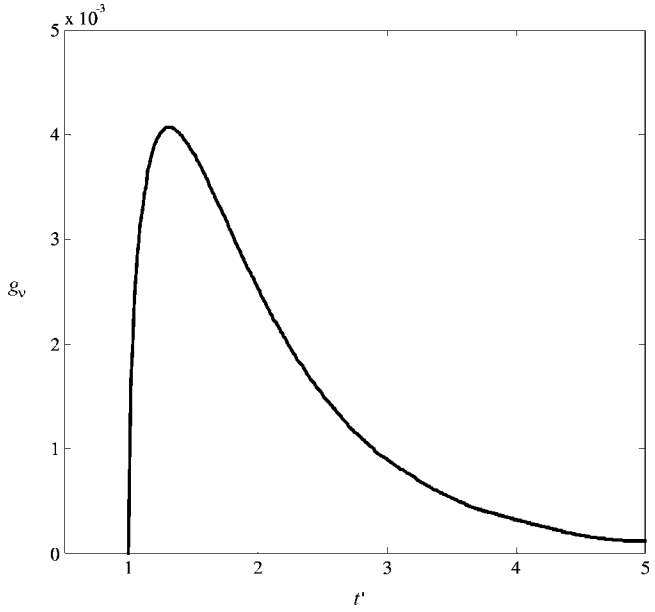


FIG. 2. Dimensionless viscoelastic Green's function in Eq. (23) is represented vs the dimensionless time $t' = t/t_r = ct/r$. The ratios t_r/τ_0 and μ_1/μ_0 are set to 1 and 9, respectively.

$$\lim_{\epsilon\to 0}\int_0^{2\pi} \frac{s_r - s_l + \epsilon\exp(i\theta)}{\mu_0}\exp\left(-\tau[-s_r - \epsilon\exp(i\theta)] \times \sqrt{\frac{s_r - s_l}{\epsilon\exp(i\theta)} + 1} + [s_r + \epsilon\exp(i\theta)]t\right) id\theta = 0, \quad (27)$$

as the convergence is governed by a term proportional to $\tau s_r/\sqrt{\epsilon}(\tau \geqslant r/c)$ in the argument of the exponential function, where we have used Eq. (21).

It is clear that Eq. (23) verifies the causality principle through the expression of the Heaviside function. It is also a good representation for numerical calculations purpose especially in the wake of the wave front since then the integration interval on τ' is very small as $t \to t_r$. Additionally, we would like to stress the fact that it does not represent a convolution on time with the elastic Green's function. Indeed, it is easy to show that such a representation is impossible.

As a numerical illustration the dispersion effect is visualized in Fig. 2, with the corresponding dimensionless elastic Green's function shown in Fig. 3. Because of the singular nature close to one in the outer integral, the numerical integration is performed by means of the Gauss quadrature method with Chebyshev polynomials. The convergence was stopped with an absolute tolerance error of one ten thousand.

It is noted that the expression of $s_{l,r}'$ in Eq. (25) is in agreement with the phenomenon of dispersion as classically encountered in wave propagation since it shows that the dispersion increases with the distance of propagation r.

VI. CONCLUSION

In this paper, a full resolution by means of Laplace transform of transient viscoelastic wave equation was proposed. The method of resolution finds roots in Strick's approach as an extension of Cagniard's method. Although our proposed method was illustrated by solving a very simple

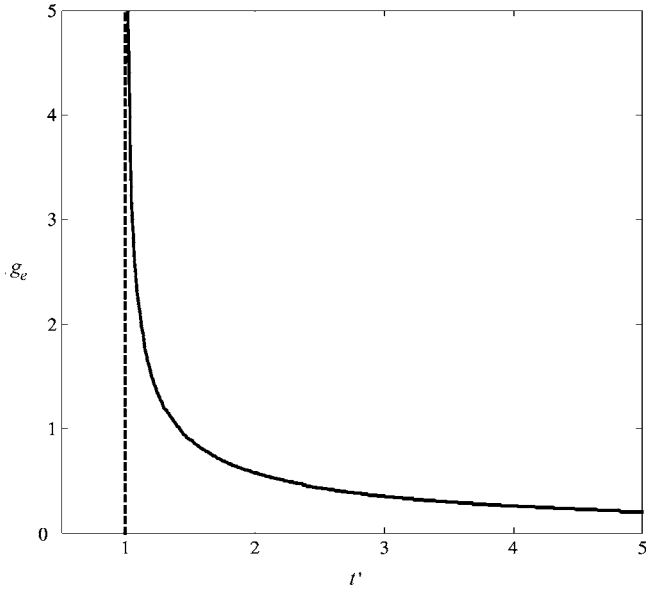


FIG. 3. Dimensionless elastic Green's function $g_e(r,t)=H(t'-1)/\sqrt{t'^2-1}$ is represented vs the dimensionless time $t'=t/t_r=ct/r$.

canonical problem, solution for the unbounded space, nonetheless, has some relevance for the implementation of numerical calculations in boundary element methods mixed with the finite element method in wave propagation and fracture mechanics.

ACKNOWLEDGMENTS

The author would like to thank the referees for their helpful suggestions.

[1]B. A. Auld, *Acoustic Fields and Waves in Solids*, 2 Vols. (Krieger Publishing, Malabon, Florida, 1990).
[2]A. Ben-Menahem and S. J. Singh, *Seismic Waves and Sources* (Springer-Verlag, New York, 1981).
[3]L. B. Freund, *Dynamic Fracture Mechanics* (Cambridge University Press, UK, 1998).
[4]K. B. Broberg, *Cracks and Fracture* (Academic Press, London, 1999).
[5]Y. A. Antipov and J. R. Willis, "Transient loading of a rapidly advancing Mode-II crack in a viscoelastic medium," Mech. Mater. **35**(3–6), 415–431 (2003).
[6]B. V. Kostrov and S. Das, *Principles of Earthquake Source Mechanics* (Cambridge University Press, UK, 1988).
[7]N. Kato, "Repeating slip events at a circular asperity: Numerical simulation with a rate-and State-dependent Friction Law," Bull. Earthquake Res. Inst., Univ. Tokyo **78**, 151–166 (2003).
[8]S. Hao, W. K. Liu, P. A. Klein, and A. J. Rosakis, "Modelling and simulation of intersonic crack growth," Int. J. Solids Struct. **41**(7), 1773–1799 (2004).
[9]A. J. Rosakis, O. Samudrala, and D. Coker, "Cracks faster than the shear wave speed," Nature (London) **284**, 1337–1340 (1999).
[10]J. D. Achenbach, *Wave Propagation in Elastic Solids* (Amsterdam, North-Holland, 1973).
[11]H. D. McNiven and Y. Mengi, "Propagation of transient cylindrical waves in an infinite viscoelastic body," Int. J. Solids Struct. **7**, 979–992 (1971).
[12]F. Mainardi and G. Turchetti, "Wave front expansions for transient viscoelastic waves," Mech. Res. Commun. **2**, 107–112 (1975).
[13]A. T. De Hoop, "Modification of Cagniard's method for solving seismic pulse problems," Appl. Sci. Res., Sect. B **8**, 349–356 (1960).
[14]D. Olsen, "Transient waves in two semi-infinite viscoelastic media separated by a plane interface," Int. J. Eng. Sci. **12**, 691–712 (1974).
[15]M. D. Verweij, "Space-time domain acoustic wave fields in media with attenuation: An analysis based on the symbolic manipulation approach," J. Acoust. Soc. Am. **95**(5), 2960–2974 (1994).
[16]R. Donnelly and R. W. Ziolkowski, "A method for constructing solutions of homogeneous partial differential equations: localized waves," Proc. R. Soc. London, Ser. A **437**, 673–692 (1992).
[17]N. V. Sushilov and R. S. C. Cobbold, "Wave propagation in media whose attenuation is proportional to frequency," Wave Motion **38**, 207–219 (2003).
[18]Z. E. A. Fellah, M. Fellah, W. Lauriks, C. Depollier, J. -Y. Chapelon, and Y. C. Angel, "Solution in time domain of ultrasonic propagation equation in a porous material," Wave Motion **38**, 151–163 (2003).
[19]P. W. Buchen, "Reflection, transmission and diffraction of SH-waves in linear viscoelastic solids," Geophys. J. R. Astron. Soc. **25**, 97–113 (1971).
[20]R. L. Ludwig and P. L. Levin, "Analytical and numerical treatment of pulsed wave propagation into a viscous fluid," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**(4), 789–792 (1995).
[21]D. V. Widder, *The Laplace Transform* (Princeton University Press, Princeton, 1941).
[22]K. Aki and P. G. Richards, *Quantitative Seismology* (University Science Books, Sausalito, California, 2002), pp. 189–195.
[23]L. Cagniard, *Reflection and Refraction of Progressive Seismics Waves*, translated by E. A. Flinn and C. H. Dix (New-York, McGraw-Hill, 1962).
[24]E. Strick, "Propagation of elastic wave motion from an impulsive source along a fluid/solid interface. II. Theoretical pressure pulse," Philos. Trans. R. Soc. London, Ser. A **251**, 465–523 (1959).

A two-dimensional model of a directional microphone: Calculation of the normal force and moment on the diaphragm

Dorel Homentcovschi,[a] Matthew J. Aubrey, and Ronald N. Miles
Department of Mechanical Engineering, SUNY Binghamton, Binghamton, New York 13902-6000

(Received 30 March 2004; accepted 11 November 2005)

It has been shown that the parasitoid fly *Ormia Ochracea* exhibits exceptional sound localization ability achieved through the mechanical coupling of its eardrums [R. N. Miles *et al.*, J. Acoust. Soc. Am. **98**, 3059–3070 (1995)]. Based on this biological system a new directional microphone has been designed, having as a basic element a special diaphragm undergoing a rocking motion. This paper considers a 2D model of the microphone in which the diaphragm is considered as a 2D plate having slits on the sides. The slits lead to a backing volume limited by an infinite rigid wall parallel to the diaphragm in its neutral position. The reflection and diffraction of an incoming plane wave by this system are studied to determine the resultant force and resultant moment of pressure upon the diaphragm. The results show that such a microphone will be driven better in the case of narrow slits and deep cavities. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149838]

PACS number(s): 43.20.El [MO] Pages: 756–768

I. INTRODUCTION

The analysis of the auditory system of the parasitoid fly *Ormia Ochracea* revealed a remarkable ability to detect the direction of the incoming sound despite the very small distance between auditory organs.[1] It was determined the fly has a special structure of the auditory system consisting of two closely spaced eardrums with a semirigid bridge connecting them. The mechanical connection between the ears causes them to move in opposite directions in response to the difference in pressure on their exterior surfaces.

Inspired by this biological system and taking advantage of modern MEMS technology Miles *et al.*,[2,3] proposed a new directional microphone integrated on a very small area. The device consists of a polysilicon diaphragm and a backplate to enable capacitive sensing of the diaphragm's motion. The diaphragm is designed to respond like a rigid plate that rocks about a central hinge. Pressure gradients on its exterior result in a net moment about the hinge and cause it to rotate. This rotation is similar to the out-of-phase motions observed in the acoustic response of the fly's ears.

In this paper we consider a simplified 2D model of the acoustic forces on this directional microphone for obtaining information concerning the parameters necessary for design purposes. In Sec. II the geometry of the model is presented, along with the corresponding PDE and the boundary conditions. Thus, the line $A'A$ in Fig. 1(b), is a plane section of the diaphragm in Fig. 1(a), undergoing a rocking motion around the axis Oy. The half-infinite lines $D'B'$ and BD are immobile parts and the segment $E'E$ corresponds to the microphones' backplate. The working domain consists of the upper half-plane $\mathcal{D}^+ \equiv \{z>0\}$, the strip $\mathcal{D}^- \equiv \{-h<z<0\}$ and the connecting slits $\mathcal{S} \equiv \{z=0; a<|x|<b\}$. Also, due to the plane-parallel geometry of the domain it is possible to reduce the boundary-value problem (BVP) to a 2D PDE even in the case of a general (plane) incoming wave. After considering this geometrical model of the problem the approach involved is that of general linear acoustics. References for these problems can be found in the classical book by Morse and Ingard[4] (also see Ref. 5). Also, most of the traditional and modern results are collected in the excellent book by Mechel *et al.*[6] The treatment in this paper is mostly analytical. However, for obtaining some results for the analyzed structure in the end some numerical computations are required.[7]

In Sec. III some representation formulas for pressure in the two domains $\mathcal{D}^+$, $\mathcal{D}^-$ are obtained. Next, a Fourier transform with respect to the x-variable was considered and the boundary conditions on the hard surfaces and the condition at infinity were applied. For the acoustical domain of frequencies that we are interested in, the solution consists of a propagating mode and an infinite number of evanescent modes. By imposing the connecting conditions along the slits $\mathcal{S}$ the results of the basic equation of the problem (Sec. IV) can be written as an integral equation. A uniqueness theorem is proven to the solution to this integral equation. Afterwards, the solution is decomposed into an odd and an even part, each of them satisfying a different integral equation. The main part of the kernels of these equations are separated and the regular parts are written in a form suitable to numerical computation. In Sec. V the two integral equations (corresponding to the odd and even parts of the solution) are reduced to two infinite systems of linear equations. This is achieved by using the spectral relationships for the two operators corresponding to the main parts of the kernels. Since the assumed form for the solution takes into consideration the proper behavior of the solution at the points $\pm a$ and $\pm b$, namely square-root singularities, the resulting infinite systems of linear equations have good convergence properties. Formulas are given for computing the resultant force and the resultant moment acting upon the diaphragm in terms of solutions of the infinite systems.

[a]Permanent address: University Politehnica of Bucharest, Applied Science Department and Institute of Mathematical Statistics and Applied Mathematics of Romanian Academy, Calea 13 Septembrie No. 13, RO-76100, Bucharest, Romania. Electronic mail: homentco@binghamton.edu

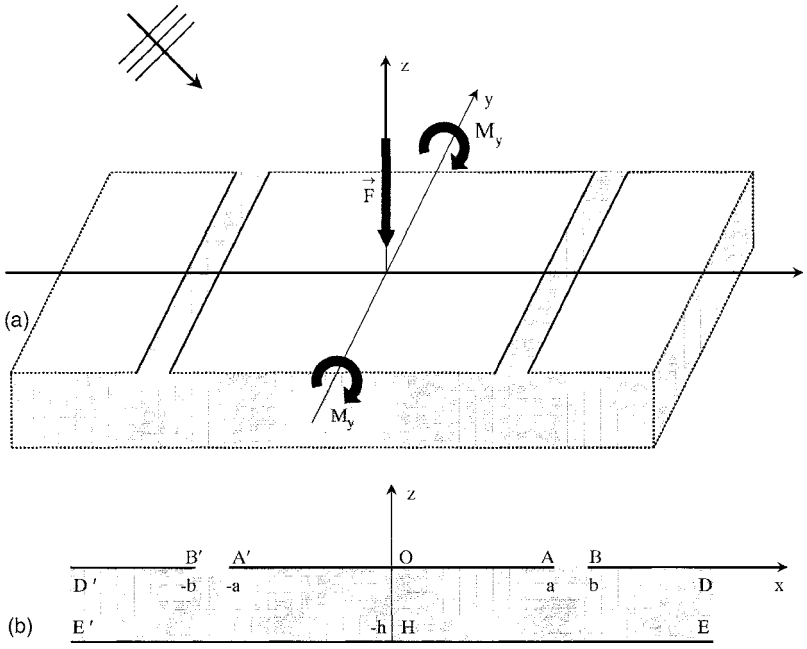


FIG. 1. (a) A sketch of the directional microphone. (b) The 2D model of the directional microphone.

A numerical analysis of the infinite linear systems is given in Sec. VI. It is based on some properties of elliptical functions, a Gauss-Legendre integration formula, and repeated use of the discrete cosine Fourier transform (DCFT). Finally, Sec. VII contains some numerical results. The graph in Fig. 4 gives the total moment of pressure as a function of the slit's width and the microphone's depth. In Figs. 5 and 6 is plotted the force's amplitude and phase delay as a function of the same parameters for a particular value of frequency. The dependence of the moment and force amplitude with frequency is plotted in Figs. 7 and 8.

The conclusion is that a directional microphone built on the ideas in Refs. 1 and 2 is driven better for very small width of the slits and quite deep back-chambers as compared with the diaphragms' width.

II. FORMULATION OF THE PROBLEM

A. The geometry of the model

In order to study the influence of reflection and diffraction of pressure waves by the edges of the diaphragm, on the diaphragm resultant force, and resultant moment we consider the model in Fig. 1. Thus, we assume a plane-parallel geometry in the direction of the Oy-axis. The segment AA' corresponds to the microphone diaphragm, AB and $B'A'$ are the two slits, and EE' is the bottom wall of the die. The origin of the Cartesian system of coordinates has been chosen at the center of the plane-parallel diaphragm and Oz axis on the upward normal direction to the diaphragm plane. We denote also by $\mathcal{D}^+$ the upper domain (the half-plane $z>0$) and by $\mathcal{D}^-$ the strip $-h<z<0$ in the lower half plane. The domain we have to study the motion of the acoustic waves is $\mathcal{D}=\mathcal{D}^+\cup\mathcal{D}^-\cup S$, where $S=B'A'\cup AB$.

B. The PDE of the problem

In the case of a harmonic motion with respect to time (of ω-angular velocity) we write the perturbation of the pressure p' as

$$p'=p_\omega(x,y,z)e^{-i\omega t}.$$

In this case the scalar wave equation for the pressure becomes

$$\frac{\partial^2 p_\omega}{\partial x^2}+\frac{\partial^2 p_\omega}{\partial y^2}+\frac{\partial^2 p_\omega}{\partial z^2}+\frac{\omega^2}{c_0^2}p_\omega=0, \tag{1}$$

which is the well known Helmholtz's equation. Here, c_0 is the unperturbed isentropic velocity of the sound.

C. Boundary conditions and condition at infinity

All of the walls are considered to be hard surfaces. Consequently, the normal velocity along the walls will be zero and the following, Neumann-type boundary condition, valid along all of these surfaces, is obtained

$$\frac{\partial p_\omega}{\partial n}=0, \tag{2}$$

n designates the direction of the normal to the surface.

Concerning the condition at infinity, the system is assumed to be under the action of an incoming plane wave described by the incident-wave direction given by angles θ_i, φ_i of Fig. 2. Thus we have

$$p_\omega^i=\frac{1}{2}\exp\left\{i\frac{\omega}{c_0}[x\sin\theta_i\cos\varphi_i+y\sin\theta_i\sin\varphi_i-z\cos\theta_i]\right\}. \tag{3}$$

In the case that the domain is $\mathcal{D}^+$ and the boundary is the whole hard plane $z=0$, it can be checked directly that the

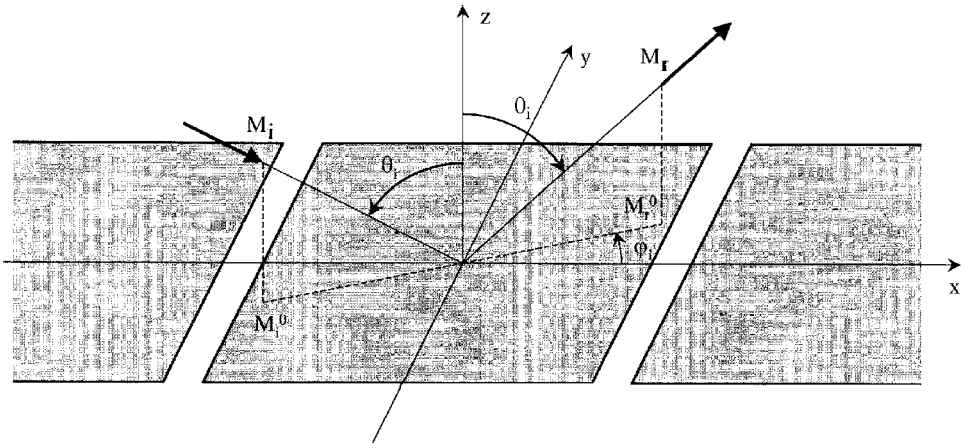


FIG. 2. The incoming and the reflected plane waves.

solution of the problem can be written by adding to p^i_ω, the reflected wave in the form

$$p^0_\omega = \frac{1}{2}\exp\left\{ i\frac{\omega}{c_0}[x\sin\theta_i\cos\varphi_i + y\sin\theta_i\sin\varphi_i - z\cos\theta_i]\right\} + \frac{1}{2}\exp\left\{ i\frac{\omega}{c_0}[x\sin\theta_i\cos\varphi_i + y\sin\theta_i\sin\varphi_i + z\cos\theta_i]\right\}. \tag{4}$$

This formula will be taken as the expression of the solution at infinity (for $z\to\infty$).

In addition, the Sommerfeld Radiation Condition is imposed, so that all of the other propagating perturbations describe outgoing waves.

D. The reduced (2D) PDE

Due to the special geometry of the problem and special form of the condition at infinity, the unknown function $p_\omega(x,y,z)$ shall be written in the form[4]

$$p_\omega(x,y,z) = p(x,z)\exp\left\{ i\frac{\omega}{c_0} y\sin\theta_i\sin\varphi_i\right\}. \tag{5}$$

The new unknown function $p(x,z)$ satisfies the 2D Helmholtz equation

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} + k_0^2 p = 0, \tag{6}$$

where the reduced wave number k_0 has the expression

$$k_0^2 \equiv \frac{\omega_0^2}{c_0^2} = \frac{\omega^2}{c_0^2}(1-\sin^2\theta_i\sin^2\varphi_i). \tag{7}$$

The solution p^0_ω can also be written as

$$p^0(x,z) = \frac{1}{2}\exp\{ik_0[x\sin\theta_0 - z\cos\theta_0]\} + \frac{1}{2}\exp\{ik_0[x\sin\theta_0 + z\cos\theta_0]\}, \tag{8}$$

where

$$k_0\sin\theta_0 = \frac{\omega}{c_0}\sin\theta_i\cos\varphi_i,$$

$$k_0\cos\theta_0 = \frac{\omega}{c_0}\cos\theta_i. \tag{9}$$

III. REPRESENTATION FORMULAS

In order to obtain some representation formulas for the function $p(x,z)$ we write

$$p(x,z) = p^0(x,z) + p^+(x,z), \quad \text{in } \mathcal{D}^+,$$

$$p(x,z) = p^-(x,z), \quad \text{in } \mathcal{D}^-. \tag{10}$$

Both functions $p^+(x,z)$, $p^-(x,z)$ are solutions of the 2D Helmholtz equation (1), satisfying the homogeneous Neumann condition $\partial p/\partial z = 0$ along the walls; on the two slits we can write

$$p^0(x,0) + p^+(x,0) = p^-(x,0), \quad \text{for } x\in S\equiv(-b,-a)\cup(a,b),$$

$$\partial p^+(x,0)/\partial z = \partial p^-(x,0)/\partial z \equiv f(x), \quad \text{for } x\in(-\infty,+\infty), \tag{11}$$

where $f(x)$ vanishes outside the slits and is an unknown function for $x\in S\equiv(-b,-a)\cup(a,b)$. In fact, as all the representation formulas will involve this function, it will be the main unknown function of the problem. This way, the unknown scattered pressures over the open slits are affected directly only by the field arriving from the other parts of the compound diffractor.

A. The solution of the 2D Helmholtz equation in $\mathcal{D}^+$

For determining an expression for the function $p^+(x,z)$ we consider a Fourier transform with respect to x,

$$P^+(\alpha,z) = \int_{-\infty}^{+\infty} p^+(x,z)e^{-i\alpha x}dx.$$

Equation (1) yields the differential equation

$$\frac{d^2P^+(\alpha,z)}{dz^2} - (\alpha^2 - k_0^2)P^+(\alpha,z) = 0. \tag{12}$$

he solution of this equation, which is vanishing at infinity upward, can be written as

$$P^+(\alpha,z) = -\frac{F(\alpha)}{\sqrt{\alpha^2 - k_0^2}} \exp\{-\sqrt{\alpha^2 - k_0^2}z\}. \tag{13}$$

The Fourier transform of the function $f(x)$ is denoted by $F(\alpha)$. Now the function $p^+(x,z)$ can be written by using the convolution theorem

$$p^+(x,z) = -\int_{-\infty}^{+\infty} f(x')K^+(x-x',z)dx', \tag{14}$$

where

$$K^+(x,z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\alpha^2 - k_0^2}} \exp\{-\sqrt{\alpha^2 - k_0^2}z\}e^{i\alpha x}d\alpha$$

$$= \frac{i}{2}H_0^1(k_0\sqrt{x^2+z^2}). \tag{15}$$

Here $H_0^1 = J_0 + iY_0$ is the Hankel function of the first kind of order zero.

B. The solution of the Helmholtz equation in $\mathcal{D}^-$

The solution of Eq. (12) satisfying the homogeneous Neumann boundary condition along the bottom surface and the second condition (11) along the surface $z=0$ can be written as

$$P^-(\alpha,z) = \frac{F(\alpha)}{\sqrt{\alpha^2 - k_0^2}} \frac{\cosh[\sqrt{\alpha^2 - k_0^2}(z+h)]}{\sinh[\sqrt{\alpha^2 - k_0^2}h]}. \tag{16}$$

Hence, the representation formula in $\mathcal{D}^-$ is obtained

$$p^-(x,z) = -\int_{-\infty}^{+\infty} f(x')K^-(x-x',z)dx', \tag{17}$$

where

$$K^-(x,z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\alpha^2 - k_0^2}} \frac{\cosh[\sqrt{\alpha^2 - k_0^2}(z+h)]}{\sinh[\sqrt{\alpha^2 - k_0^2}h]} e^{i\alpha x}d\alpha. \tag{18}$$

This function can also be written as

$$K^-(x,z) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{1}{\sqrt{\alpha^2 - t_0^2}} \frac{\cosh\left[\sqrt{\alpha^2 - t_0^2}\left(1 + \frac{z}{h}\right)\right]}{\sinh[\sqrt{\alpha^2 - t_0^2}]} \times \exp\left\{i\alpha\frac{x}{h}\right\}d\alpha, \tag{19}$$

where

$$t_0 = k_0 h = \frac{\omega_0}{c_0}h.$$

The inverse Fourier transform in (19) cannot be performed in finite form. However, the residue theorem can be applied in order to evaluate this integral. There are in this case some real poles which describe undamped waves, corresponding to resonance frequencies, and also some imaginary poles which give evanescent modes. The Sommerfeld condition requires that only the real positive poles which are providing outgoing waves have to be considered. Hence the integration contour will be that drawn in Fig. 3. A simple discussion about the application of the Fourier transform for solving the wave equation can be found in Ref. 5, pp. 293–295.

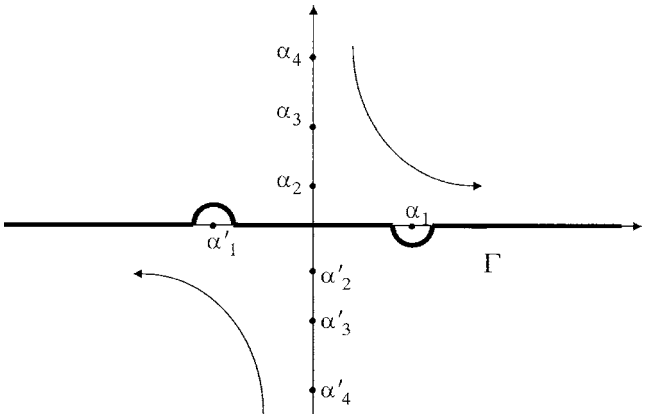


FIG. 3. The integration path Γ.

To be specific, for the range of parameters of this particular problem ($\omega < 2\pi \cdot 20$ kHz) the first pole at the point $\alpha_1 = t_0 (t_0 > 0)$ is real. The next one, $\alpha_2 = \sqrt{t_0^2 - \pi^2}$, is imaginary as well as all the other poles. Consequently the function $K^-(x,y)$ can be written as

$$K^-(x,z) = i\frac{\exp\{ik_0|x|\}}{2t_0} + \sum_{n=1}^{\infty}(-1)^n \frac{\cos[n\pi(1+z/h)}{\sqrt{n^2\pi^2 - t_0^2}} \times \exp\left\{-\sqrt{n^2\pi^2 - t_0^2}\frac{|x|}{h}\right\}. \tag{20}$$

Now, once the function $f(x)$ is determined, the formulas (14) and (17) can be used for obtaining the pressure field in any point of the domain $\mathcal{D}^-$.

Remark 1: In the case that the working frequency ω is increasing, the poles on the upper imaginary axis are moving toward the real positive semiaxis providing more propagating modes corresponding to different resonance frequencies. Hence, the method developed in this paper can be applied to all (finite) working frequencies if the supplementary propagating modes are taken into consideration.

IV. THE BASIC EQUATION

For obtaining the equation to determine the function $f(x)$, the representation formulas (17) and (20) could be used. Since the formula corresponding to the domain $\mathcal{D}^-$ involves an infinite series it is preferable to work with closed expressions in the Fourier transform plane. Then, the first condition (11) yields the basic equation in the form

$$P_S A(D) f(x) = g(x), \quad x \in S,$$

$$f(x)=0, \quad x \in R-S, \tag{21}$$

where

$$g(x)=\exp\{ik_0 x \sin\theta_0\}.$$

Here P_S is the restriction operator to the reunion of intervals S, and $A(D)$ is a pseudodifferential operator with the symbol

$$A(\alpha)=\frac{1}{\sqrt{\alpha^2-t_0^2}}\frac{\cosh\sqrt{\alpha^2-t_0^2}}{\sinh\sqrt{\alpha^2-t_0^2}}+\frac{1}{\sqrt{\alpha^2-t_0^2}}.$$

The pseudodifferential operator acts on the function $f(x)$ as

$$A(D)f(x)=\frac{1}{2\pi}\int_\Gamma A(\alpha)F(\alpha)\exp\left\{i\alpha\frac{x}{h}\right\}d\alpha.$$

The contour Γ, symmetrical with respect to origin, is shown in Fig. 3.

Thus, for the function $f(x)$ the pseudodifferential equation (21) is obtained. The contour Γ assures that the solution satisfies Sommerfeld's condition at infinity.

Alternatively, the basic equation can be written as the integral equation

$$\int_S f(x')K(x-x')dx'=g(x), \quad x \in S \equiv (-b,-a)\cup(a,b), \tag{22}$$

where the kernel K is

$$K(x)=\frac{1}{2\pi}\int_\Gamma\left[\frac{1}{\sqrt{\alpha^2-t_0^2}}\frac{\cosh\sqrt{\alpha^2-t_0^2}}{\sinh\sqrt{\alpha^2-t_0^2}}+\frac{1}{\sqrt{\alpha^2-t_0^2}}\right]\exp\left\{i\alpha\frac{x}{h}\right\}d\alpha. \tag{23}$$

A. The uniqueness theorem

It shall be proven that the basic equation has at most one solution. Indeed, the homogeneous equation [corresponding to $g(x)\equiv 0$] can be written as

$$\int_S f(x')K(x-x')dx'=0, \quad x \in S.$$

Multiplying by $\overline{f(x)}$ (the overbar denotes the complex conjugate function) and integrating along S there results

$$\int\int K(x-x')f(x')\overline{f(x)}dx'dx=0,$$

the integral being taken over the whole Oxx'-plane. By substituting the expression (23) of the kernel there results

$$\int_\Gamma\left[\frac{1}{\sqrt{\alpha^2-t_0^2}}\frac{\cosh\sqrt{\alpha^2-t_0^2}}{\sinh\sqrt{\alpha^2-t_0^2}}+\frac{1}{\sqrt{\alpha^2-t_0^2}}\right]F(\alpha)\overline{F(\bar{\alpha})}d\alpha=0.$$

Using the residue theorem, in the form valid for the case the contour remains along the real axis, the following expression is obtained:

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{\alpha^2-t_0^2}}\frac{\cosh\sqrt{\alpha^2-t_0^2}}{\sinh\sqrt{\alpha^2-t_0^2}}|F(\alpha)|^2d\alpha+\int_{|\alpha|>t_0}\frac{|F(\alpha)|^2}{\sqrt{\alpha^2-t_0^2}}d\alpha+i\left\{\int_{-t_0}^{+t_0}\frac{|F(\alpha)|^2}{\sqrt{t_0^2-\alpha^2}}d\alpha+\pi\frac{|F(-t_0)|^2+|F(t_0)|^2}{2t_0}\right\}=0.$$

The imaginary part in this relation gives

$$F(t_0)=F(-t_0)=0,$$

$$\int_{-t_0}^{+t_0}\frac{|F(\alpha)|^2}{\sqrt{t_0^2-\alpha^2}}d\alpha=0.$$

The last relationship yields $F(\alpha)=0$ almost everywhere in the interval $(-t_0,t_0)$. As the Fourier transform of a summable function over a finite interval is analytic in the whole plane there results $F(\alpha)=0$ all over; its inverse Fourier transform $f(x)$ is also vanishing along the whole real axis. We in fact have proved the uniqueness theorem.

Theorem 2: *Equation (22) has at most a solution in the space of q-summable functions* L_q, $(1<q\leqslant 2)$.

B. Odd and even solutions of the basic equation

The symmetry of the integration intervals with respect to the origin makes it possible to write particular integral equations for the odd and even part of the solution. It is evident from formula (23) that $K(x)=K(-x)$ which is in fact a physical embodiment of reciprocity. Consequently, the kernel is a function of $|x|$ that will be denoted also by $K(|x|)$. We introduce now the odd and even part of the solutions of the basic integral equation by

$$f_o(x)=0.5[f(x)-f(-x)],$$

$$f_e(x)=0.5[f(x)+f(-x)],$$

or, equivalently,

$$f(x)=f_e(x)+f_o(x),$$

$$f(-x)=f_e(x)-f_o(x).$$

Hence, Eq. (22) gives

$$\int_a^b K(|x-x'|)[f_e(x')+f_o(x')]dx'+\int_a^b K(|x+x'|)[f_e(x')-f_o(x')]dx'=g(x), \quad x \in (a,b), \tag{24}$$

$$\int_a^b K(|x+x'|)[f_e(x')+f_o(x')]dx'+\int_a^b K(|x-x'|)[f_e(x')-f_o(x')]dx'=g(-x), \quad x \in (a,b). \tag{25}$$

Now, from the sum and the difference of Eqs. (24) and (25) we obtain

$$\int_a^b K^{(o)}(x,x')f_o(x')=\frac{\pi}{2}g_o(x'), \quad x \in (a,b), \tag{26}$$

$$\int_a^b K^{(e)}(x,x')f_e(x') = \frac{\pi}{2} g_e(x'), \quad x \in (a,b). \tag{27}$$

These are two independent equations for determining the functions $f_o(x)$, $f_e(x)$. We have denoted

$$K^{(o)}(x,x') = \frac{\pi}{2}[K(|x-x'|) - K(|x+x'|)],$$

$$K^{(e)}(x,x') = \frac{\pi}{2}[K(|x-x'|) + K(|x+x'|)],$$

$$g_o(x) \equiv 0.5[g(x) - g(-x)] = i \sin(k_0 x \sin\theta_0),$$

$$g_e(x) \equiv 0.5[g(x) + g(-x)] = \cos(k_0 x \sin\theta_0).$$

C. Analysis of the of integral equation kernels

An analysis of the kernels of the integral equations shall now be produced. Thus, the main (singular) part of the kernel, which determines the behavior (singularity) of the solution at the intervals ends, will be separated. Also, the kernel will be written in a form directly amenable to efficient numerical methods.

Thus,

$$K^+(x) \equiv \frac{1}{2\pi}\int_\Gamma \frac{e^{ikx}}{\sqrt{k^2-k_0^2}}dk = -\frac{1}{2}Y_0(k_0|x|) + \frac{i}{2}J_0(k_0|x|),$$

$$K^-(x) \equiv \frac{1}{2\pi}\int_\Gamma \frac{\coth\sqrt{t^2-t_0^2}}{\sqrt{t^2-t_0^2}} e^{itx/h}dt$$
$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\left(\frac{\coth t_0\sqrt{t^2-1}}{\sqrt{t^2-1}} - \frac{1}{t_0(t^2-1)}\right)e^{ik_0tx}dt$$
$$+ \frac{i}{2t_0}e^{ik_0|x|}.$$

Hence,

$$K^-(x) = \frac{1}{\pi}\int_0^1 \left\{\frac{1}{t_0(1-t^2)} - \frac{\cot(t_0\sqrt{1-t^2})}{\sqrt{1-t^2}}\right\}\cos(k_0tx)dt$$
$$+ \frac{1}{\pi}\int_1^2 \left\{\frac{\coth(t_0\sqrt{t^2-1})}{\sqrt{t^2-1}} - \frac{1}{t_0(t^2-1)} - \frac{1}{\sqrt{t^2-1}}\right\}$$
$$\times\cos(k_0tx)dt + \frac{1}{\pi}\int_2^\infty \left\{\frac{\coth(t_0\sqrt{t^2-1})}{\sqrt{t^2-1}} - \frac{1}{\sqrt{t^2-1}}\right\}$$
$$\times\cos(k_0tx)dt + \frac{1}{\pi t_0}\int_0^2 \frac{\cos(k_0tx)}{t^2-1}dt + i\frac{\cos(k_0|x|)}{2t_0}$$
$$- \frac{1}{2}Y_0(k_0|x|),$$

the last integral is being considered as a Cauchy principal-value integral. In obtaining this formula we have used also the integral

$$\int_0^\infty \frac{\cos(k_0tx)}{t^2-1}dt = -\frac{\pi}{2}\sin(k_0|x|),$$

which can be found in Ref. 8.

We have

$$K(x) = K^+(x) + K^-(x),$$

and hence, by some changes of variables

$$K(x) = \frac{1}{\pi}\int_0^{\pi/2}\left\{\frac{1}{t_0\sin u} - \cot(t_0\sin u)\right\}\cos(k_0x\cos u)du$$
$$+ \frac{1}{\pi}\int_0^d \left\{\coth(t_0\sinh u) - \frac{1}{t_0\sinh u} - 1\right\}$$
$$\times\cos(k_0x\cosh u)du + \frac{1}{\pi}\int_d^\infty \{\coth(t_0\sinh u) - 1\}$$
$$\times\cos(k_0x\cosh u)du + \frac{1}{\pi t_0}\int_0^2 \frac{\cos(k_0xt)}{t^2-1}dt$$
$$+ i\frac{\cos(k_0|x|)}{2t_0} - Y_0(k_0|x|) + \frac{i}{2}J_0(k_0|x|),$$

where $d=\cosh^{-1}(2)$.

Finally the kernel of Eqs. (26) and (27) can be written as

$$K^{(o)}(x,x') = -\ln\frac{|x-x'|}{x+x'} + \frac{1}{t_0}\int_0^2 \frac{\sin(k_0xt)\sin(k_0x't)}{t^2-1}dt + \frac{i\pi}{4}\{J_0(k_0|x-x'|) - J_0(k_0|x+x'|)\}$$
$$+ \frac{i\pi}{2t_0}\sin(k_0x)\sin(k_0x') - \frac{\pi}{2}\left\{Y_0(k_0|x-x'|) - Y_0(k_0|x+x'|) - \frac{2}{\pi}\ln\frac{|x-x'|}{x+x'}\right\}$$
$$+ \int_0^{\pi/2}\left\{\frac{1}{t_0\sin u} - \cot(t_0\sin u)\right\}\sin(k_0x\cos u)\sin(k_0x'\cos u)du$$
$$+ \int_0^d \left\{\coth(t_0\sinh u) - \frac{1}{t_0\sinh u} - 1\right\}\sin(k_0x\cosh u)\sin(k_0x'\cosh u)du$$
$$+ \int_d^\infty \{\coth(t_0\sinh u) - 1\}\sin(k_0x\cosh u)\sin(k_0x'\cosh u)du,$$

$$K^{(e)}(x,x') = -\ln|x^2 - x'^2| + \frac{1}{t_0}\int_0^2 \frac{\cos(k_0 x t)\cos(k_0 x' t)}{t^2-1}dt + \frac{i\pi}{4}\{J_0(k_0|x-x'|) + J_0(k_0|x+x'|)\}$$
$$+ \frac{i\pi}{2t_0}\cos(k_0 x)\cos(k_0 x') - \frac{\pi}{2}\left\{Y_0(k_0|x-x'|) + Y_0(k_0|x+x'|) - \frac{2}{\pi}\ln|x^2-x'^2|\right\}$$
$$+ \int_0^{\pi/2}\left\{\frac{1}{t_0\sin u} - \cot(t_0\sin u)\right\}\cos(k_0 x\cos u)\cos(k_0 x'\cos u)du$$
$$+ \int_0^{d}\left\{\coth(t_0\sinh u) - \frac{1}{t_0\sinh u} - 1\right\}\cos(k_0 x\cosh u)\cos(k_0 x'\cosh u)du$$
$$+ \int_d^{\infty}\{\coth(t_0\sinh u) - 1\}\cos(k_0 x\cosh u)\cos(k_0 x'\cosh u)du.$$

The first term on the rhs proves that the integral equation has a logarithmic singularity. The first integral is a Cauchy principal-value integral and all the other terms are regular, or regular integrals.

Finally, Eqs. (26) and (27) will be written in the form

$$-\int_a^b f_o(x')\ln\frac{|x-x'|}{x+x'}dx' + \int_a^b f_o(x')\tilde{K}^{(o)}(x,x')dx' = \frac{\pi}{2}g_o(x), \quad x\in(a,b), \tag{28}$$

$$-\int_a^b f_e(x')\ln|x-x'^2|dx' + \int_a^b f_o(x')\tilde{K}^{(e)}(x,x')dx' = \frac{\pi}{2}g_e(x), \quad x\in(a,b), \tag{29}$$

separating the main (singular) terms,

$$\tilde{K}^{(o)}(x,x') = K^{(o)}(x,x') + \ln\frac{|x-x'|}{x+x'},$$

$$\tilde{K}^{(e)}(x,x') = K^{(e)}(x,x') + \ln|x^2-x'^2|.$$

V. REDUCTION OF THE INTEGRAL EQUATIONS TO INFINITE SYSTEMS OF LINEAR EQUATIONS

A. The spectral relationships for logarithmic operators and the spaces $L^{1/2}_{2o}(a,b)$, $L^{1/2}_{2e}(a,b)$

For approaching the integral equations (28) and (29) the spectral relationships that invert the main (singular) part of the kernels of the integral equations will be used. For the "odd" case this spectral relationship was given by Aleksandrov *et al.* in Ref. 9,

$$-\int_a^b \frac{T_n(X_o')}{\sqrt{(b^2-x'^2)(x'^2-a^2)}}\ln\frac{|x-x'|}{x+x'}dx' = \mu_n^{(o)}T_n(X_o), \quad n\geqslant 0, \tag{30}$$

where

$$X_o = \cos\left\{\frac{\pi}{K'(c)}F\left(\arcsin\sqrt{\frac{b^2-x^2}{b^2-a^2}}, c'\right)\right\}$$

$$\mu_0^{(o)} = \frac{\pi}{b}K(c), \quad \mu_n^{(o)} = \frac{K'(c)}{nb}\tanh\frac{n\pi K(c)}{K'(c)}, \quad n\geqslant 1,$$

$$c = \frac{a}{b}, \quad c' = \sqrt{1-c^2}, \quad K' = K'(c) = K(c'). \tag{31}$$

$T_n(X_o) = \cos(n\arccos X_o)$ denotes the Chebyshev polynomials of the first kind. F is the elliptic integral of the first kind and by $K(c)$ denotes the complete elliptical integral of the first kind.

Relation (30) yields the formula

$$-\int_a^b\int_a^b \frac{T_n(X_o')T_m(X_o)}{\sqrt{(b^2-x'^2)(x'^2-a^2)}\sqrt{(b^2-x^2)(x^2-a^2)}} \times \ln\frac{|x-x'|}{x+x'}dxdx' = \mu_n^{(o)}\int_a^b \frac{T_n(X_o)T_m(X_o)}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx.$$

Due to the symmetry of the first term in this expression with respect to x and x' we can write

$$\mu_n^{(o)}\int_a^b \frac{T_n(X_o)T_m(X_o)}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx = \mu_m^{(o)}\int_a^b \frac{T_n(X_o)T_m(X_o)}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx.$$

As $\mu_n^{(o)} \neq \mu_m^{(o)}$ there results the orthogonality relationship

$$\int_a^b \frac{T_n(X_o)T_m(X_o)}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx = \begin{cases} 0 & \text{for } n\neq m \\ K'(c)/(2b), & \text{for } n=m\neq 0 \\ K'(c)/b, & \text{for } n=m=0. \end{cases}$$

Inspired by this formula the inner product

$$\langle f(x), g(x)\rangle_o = \int_a^b \frac{f(X_o)\overline{g(X_o)}}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx$$

is defined and, correspondingly, the norm

$$\|f(x)\|_o^2 = \int_a^b \frac{|f(X_o)|^2}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx.$$

We denote by $\mathbf{L}_{2o}^{1/2}(a,b)$ the completion of the space of continuous functions on (a,b) in this norm. It is obvious that $\mathbf{L}_{2o}^{1/2}(a,b)$ is a Hilbert space. In this space the set of functions $T_n(X_o)$ is an orthogonal complete system.

A similar analysis can be made for the main (singular) part of the "even" equation. In this case the spectral relationship is

$$-\int_a^b \frac{2x' T_n(X_e')}{\sqrt{(b^2-x'^2)(x'^2-a^2)}} \ln|x^2-x'^2|dx' = \mu_n^{(e)} T_n(X_e), \quad n \geqslant 0,$$

where

$$X_e = \frac{2}{b^2-a^2}\left[x^2 - \frac{b^2+a^2}{2}\right],$$

$$\mu_0^{(e)} = \pi \ln \frac{4}{b^2-a^2}, \quad \mu_n^{(e)} = \frac{\pi}{n}, \quad n \geqslant 1.$$

The orthogonality relationship

$$\int_a^b \frac{2xT_n(X_e)T_m(X_e)}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx = \begin{cases} 0 & \text{for } n \neq m \\ \pi/2, & \text{for } n = m \neq 0 \\ \pi, & \text{for } n = m = 0 \end{cases}$$

can be proven and also the "even" inner product

$$\langle f(x), g(x)\rangle_e = \int_a^b \frac{2xf(X_e)\overline{g(X_e)}}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx$$

may be defined along with the corresponding norm

$$\|f(x)\|_e^2 = \int_a^b \frac{2x|f(X_e)|^2}{\sqrt{(b^2-x^2)(x^2-a^2)}}dx.$$

We denote by $\mathbf{L}_{2e}^{1/2}(a,b)$ the completion of the space of continuous functions on $[a,b]$ in the $\mathbf{L}_{2e}^{1/2}(a,b)$ norm. The set of polynomials $T_n(X_e)$ is an orthogonal complete system in this Hilbert space.

B. The infinite system of linear equations equivalent to the "odd" integral equation

According to the theory developed in previous section the solution of the integral equation (28) can be written in the form

$$f_o(x) = \frac{\sum_{n=0}^{\infty} A_n^{(o)} T_n(X_o)}{\sqrt{(b^2-x^2)(x^2-a^2)}}, \tag{32}$$

where the coefficients $A_n^{(o)}$ have to be determined. Equation (28) becomes

$$\sum_{n=0}^{\infty} \mu_n^{(o)} A_n^{(o)} T_n(X_o) + \sum_{n=0}^{\infty} A_n^{(o)} \int_a^b \frac{T_n(X_o')\tilde{K}^{(o)}(x,x')dx'}{\sqrt{(b^2-x'^2)(x'^2-a^2)}} = \frac{\pi}{2} g_o(x). \tag{33}$$

In order to obtain a system of linear equations for unknown coefficients A_n consider the inner product with $T_m(X_o)$. [The spectral postmultiplication of the integral equation with the function $T_m(X_o)$ is somewhat arbitrary since one cannot hope to achieve a spectral diagonal (perfect separability). Rather one hopes that the Chebyshev polynomial function in question will strengthen the diagonal of the resulting spectral system of equations.] The relationship (33) becomes

$$\delta_m^{(o)2}\mu_m^{(o)}A_m^{(o)} + \sum_{n=0}^{\infty} \tilde{K}_{mn}^{(o)} A_n^{(o)} = g_m^{(o)}, \quad m = 0,1,2,\ldots, \tag{34}$$

where

$$\delta_m^{(o)2} = \begin{cases} K'(c)/(2b), & \text{for } m \neq 0 \\ K'(c)/b & \text{for } m = 0, \end{cases}$$

$$\tilde{K}_{mn}^{(o)} = \int_a^b \int_a^b \frac{T_m(X_o)T_n(X_o')\tilde{K}^{(o)}(x,x')dxdx'}{\sqrt{(b^2-x^2)(x^2-a^2)}\sqrt{(b^2-x'^2)(x'^2-a^2)}}, \tag{35}$$

$$g_m^{(o)} = \frac{\pi}{2}\int_a^b \frac{T_m(X_o)g_o(x)dx}{\sqrt{(b^2-x^2)(x^2-a^2)}}. \tag{36}$$

The way the infinite linear system was obtained (34) shows that the system is equivalent to the integral equation (28). Since the linear system comes directly from the integral equation the proof of uniqueness also applies to the linear system. That is the infinite linear system always has an unique solution.

Remark 3: The conversion of the relationship (33) into a linear system (34) is equivalent to a Galerkin formulation. Alternatively, a collocation method can be used by imposing Eq. (33) to be satisfied at a special set of points. In this case the Galerkin procedure is preferred due to its connection with the above defined inner products.

C. The infinite system of linear equations equivalent to the "even" integral equation

The solution of the integral equation (29) shall be written in the form

$$f_e(x) = \frac{2x\sum_{n=0}^{\infty} A_n^{(e)} T_n(X_e)}{\sqrt{(b^2-x^2)(x^2-a^2)}}, \tag{37}$$

where the coefficients $A_n^{(e)}$ have to be determined from the integral equation. We obtain

$$\sum_{n=0}^{\infty} \mu_n^{(e)} A_n^{(e)} T_n(X_e) + \sum_{n=0}^{\infty} A_n^{(e)} \int_a^b \frac{2x' T_n(X_e')\tilde{K}(x,x')dx'}{\sqrt{(b^2-x'^2)(x'^2-a^2)}} = \frac{\pi}{2} g_e(x). \tag{38}$$

The "even" inner product with $T_m(X_e)$ yields the infinite system of linear equations

$$\delta_m^{(e)2}\mu_m^{(e)}A_m^{(e)} + \sum_{n=0}^{\infty} \tilde{K}_{mn}^{(e)} A_n^{(e)} = g_m^{(e)}, \quad m = 0,1,2,\ldots, \tag{39}$$

where

$$\delta_m^{(e)2} = \begin{cases} \pi/2, & \text{for } m \neq 0 \\ \pi, & \text{for } m = 0. \end{cases}$$

$$\tilde{K}_{mn}^{(e)} = \int_a^b \int_a^b \frac{4xx' T_m(X_e) T_n(X_e') \tilde{K}^{(e)}(x,x') dx dx'}{\sqrt{(b^2 - x^2)(x^2 - a^2)}\sqrt{(b^2 - x'^2)(x'^2 - a^2)}}, \tag{40}$$

$$g_m^{(e)} = \frac{\pi}{2} \int_a^b \frac{2x T_m(X_e) g_e(x) dx}{\sqrt{(b^2 - x^2)(x^2 - a^2)}}. \tag{41}$$

D. The forces due to fluid pressure upon the diaphragm

In order to obtain the resultant force and the resultant moment of pressure upon the diaphragm consider the relationships

$$p_\omega(x,-0) = \int_S f(x') K^-(|x - x'|) dx',$$

$$p_\omega(x,+0) = p^0(x,0) - \int_S f(x') K^+(|x - x'|) dx',$$

giving the pressure on two faces of the diaphragm in terms of the function $f(x)$. Hence

$$p_\omega(x,-0) - p_\omega(x,+0) = -p^0(x,0) + \int_S f(x') K(|x - x'|) dx'.$$

By introducing the odd and even parts of the function f we can write this relationship in the form

$$p_\omega(x,-0) - p_\omega(x,+0) = -p^0(x,0) + \frac{2}{\pi} \int_a^b f_o(x') K_o(x,x') dx' + \frac{2}{\pi} \int_a^b f_e(x') K_e(x,x') dx'.$$

Hence,

$$F \equiv \int_{-a}^{+a} [p_\omega(x,-0) - p_\omega(x,+0)] dx = -\int_{-a}^{+a} p^0(x,0) dx + \frac{2}{\pi} \int_{-a}^{+a} dx \int_a^b f_e(x') K_e(x,x') dx',$$

$$M \equiv \int_{-a}^{+a} x[p_\omega(x,-0) - p_\omega(x,+0)] dx = -\int_{-a}^{+a} x p^0(x,0) dx + \frac{2}{\pi} \int_{-a}^{+a} x dx \int_a^b f_o(x') K_o(x,x') dx'.$$

By using the expressions (32) and (37) of the two functions f_o, f_e we obtain

$$F = -2 \frac{\sin(ak_0 \sin\theta_0)}{k_0 \sin\theta_0} + \sum_{m=0}^{\infty} A_m^{(e)} F_m, \tag{42}$$

$$M = 2ai \frac{\cos(ak_0 \sin\theta_0)}{k_0 \sin\theta_0} - 2i \frac{\sin(ak_0 \sin\theta_0)}{(k_0 \sin\theta_0)^2} + \sum_{m=0}^{\infty} A_m^{(o)} M_m, \tag{43}$$

where

$$F_n = \frac{4}{\pi} \int_0^a dx \int_a^b \frac{2x' T_n(X_e') K^{(e)}(x,x') dx'}{\sqrt{(b^2 - x'^2)(x'^2 - a^2)}}, \tag{44}$$

$$M_n = \frac{4}{\pi} \int_0^a x dx \int_a^b \frac{T_n(X_o') K^{(o)}(x,x') dx'}{\sqrt{(b^2 - x'^2)(x'^2 - a^2)}}. \tag{45}$$

Remark 4: We note that due to 2D model considered in this paper the pressure does not depend upon the y-variable. Correspondingly, the total normal force and moment on the diaphragm can be obtained by multiplying the quantities F and M [given by formulas (42) and (43)] by L_y, the diaphragm's width.

VI. NUMERICAL ANALYSIS OF THE LINEAR SYSTEMS

For both the "odd" and "even" cases an infinite system of linear equations for solving the problem was obtained. However, the coefficients of these systems cannot be evaluated analytically, hence they must be computed by numerical methods.

A. The "odd" case

To begin consider the problem of computation of generalized Fourier coefficients of a given continuous function with respect to the orthogonal bases $\{T_m(X_o)\}$ in the Hilbert space $\mathbf{L}_{2o}^{1/2}(a,b)$. For example,

$$g_m^{(o)} = \frac{\pi}{2} \langle g_o(x), T_m(X_o) \rangle.$$

Let $h(x)$ be a smooth function defined on the interval (a,b). Then,

$$\langle h(x), T_m(X_o) \rangle_o = \int_a^b \frac{h(x) T_m(X_o) dx}{\sqrt{(b^2 - x^2)(x^2 - a^2)}}.$$

By inverting the relationship (31) the following expression for x can be written:

$$x = \sqrt{b^2 - (b^2 - a^2)\mathrm{sn}^2\left[\frac{K'}{\pi} \arccos(X_o)\right]}, \tag{46}$$

where sn is the Jacobian elliptic function and,

$$X_o = \cos\theta. \tag{47}$$

By means of the changes of variables (46) and (47) there results

$$\int_a^b \frac{h(x) T_m(X_o) dx}{\sqrt{(b^2 - x^2)(x^2 - a^2)}} = \frac{K'}{\pi b} \int_0^\pi h\left(\sqrt{b^2 - (b^2 - a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)}\right) \cos(m\theta) d\theta. \tag{48}$$

By using the relationship[8]

$$\mathrm{sn}(u+2K',c') = -\mathrm{sn}(u,c'),$$

we can write

$$\mathrm{sn}^2\left[\frac{K'}{\pi}(\theta+2\pi),c'\right] = \mathrm{sn}^2\left(\frac{K'}{\pi}\theta+2K',c'\right) = \mathrm{sn}^2\left(\frac{K'}{\pi}\theta,c'\right).$$

Hence, the function

$$\tilde{h} \equiv h\left(\sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)}\right)$$

is a smooth, even, 2π-periodic function. The generalized Fourier coefficient can be written as

$$h_n^{(o)} \equiv \langle h(x),T_m(X_o)\rangle_o = \frac{K'}{2b}\tilde{h}_m,$$

where

$$\tilde{h}_m = \frac{1}{\pi}\int_0^{2\pi} h\left(\sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)}\right)\cos(m\theta)d\theta, \quad m=0,1,\ldots$$

are the (cosine) Fourier coefficients of the smooth function $\tilde{h}(\theta)$. They can be well approximated by means of the discrete cosine Fourier transform and can be efficiently computed by using a DCT algorithm based on FFT.[10]

For determining the coefficients $\tilde{K}_{mn}^{(o)}$, consider the expression

$$h_{mn}^{(o)} = \langle\langle h(x,x'),T_n(X_o')\rangle_o,T_m(X_o)\rangle_o = \int_a^b\int_a^b \frac{T_m(X_o)T_n(X_o')h(x,x')dxdx'}{\sqrt{(b^2-x^2)(x^2-a^2)}\sqrt{(b^2-x'^2)(x'^2-a^2)}}.$$

By using the change of variables

$$\begin{cases} x = \sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)} \\ x' = \sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta'\right)}, \end{cases}$$

there results

$$h_{mn}^{(o)} = \frac{K'^2}{4\pi^2b^2}\int_0^{2\pi}\int_0^{2\pi} h\left(\sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)},\sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta'\right)}\right)\cdot\cos(m\theta)\cos(n\theta')d\theta d\theta'.$$

Thus, in the case the function $h(x,y)$ is given by its analytical form, the coefficients $h_{mn}^{(o)}$ can be computed by means of a 2D discrete cosine (Fourier) transform.

In order to evaluate the contribution to $\tilde{K}_{mn}^{(o)}$ coefficients of the integral terms in the expression of the kernel $K^{(o)}\times(x,x')$ we introduce the functions

$$S_n^o(k_0t) = \langle\sin(k_0tx),T_n(X_o)\rangle_o$$

which can be computed as

$$S_n^o(k_0t) = \frac{K'}{2\pi b}\int_0^{2\pi}\sin\left[k_0t\sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)}\right]\cos(n\theta)d\theta$$

by using a DCT algorithm. Now due to the special form of the integrals a parallel algorithm for computing simultaneously all of the integrals corresponding to all values $0\leqslant n\leqslant N$, $0\leqslant m\leqslant n$ by certain quadrature formulas may be employed. Consider, as an example, the Cauchy principal-value integral

$$I_1^o \equiv \int_a^b\int_a^b dxdx'\int_0^2 \frac{\sin(k_0tx)T_m(X_o)}{\sqrt{(b^2-x^2)(x^2-a^2)}}\times\frac{\sin(k_0tx')T_n(X_o')}{\sqrt{(b^2-x'^2)(x'^2-a^2)}}\frac{dt}{t^2-1}.$$

We can write

$$I_1^o = \int_0^2 \frac{S_n^o(k_0t)S_m^o(k_0t)}{t^2-1}dt = \sum_{j=1}^{2p}\frac{w_j}{t_j^2-1}S_n^o(k_0t_j)S_m^o(k_0tj),$$

w_j,t_j being the weights and, respectively, the nodes of an even $(2p)$ Gauss-Legendre quadrature formula.

By the same method the moment coefficients M_n may be obtained

$$M_n = \frac{2K'}{\pi^2b}\int_0^a xdx\int_0^{2\pi} K^{(o)}\times\left(x,\sqrt{b^2-(b^2-a^2)\mathrm{sn}^2\left(\frac{K'}{\pi}\theta\right)}\right)\cos(n\theta)d\theta.$$

For the explicit terms in the kernel $K^{(o)}$ the calculation is straightforward. For the terms containing integrals take as an example the same integral as before. Thus,

$$IM_1 \equiv \frac{4}{\pi}\int_0^a xdx\int_a^b \frac{T_n(X_o')dx'}{\sqrt{(b^2-x'^2)(x'^2-a^2)}}\times\int_0^2\frac{\sin(k_0xt)\sin(k_0x't)}{t^2-1}dt = \frac{4}{\pi}\times\int_0^2\left(\frac{\sin(k_0at)}{k_0^2t^2}-a\frac{\cos(k_0at)}{k_0t}\right)\frac{S_n^o(k_0t)}{t^2-1}dt.$$

The resulting integral can also be computed by means of an even Gauss-Legendre quadrature formula.

B. The "even" case

A similar analysis can be performed for the "even" problem. In this case only the final formulas shall be given. Thus, for a smooth function $h(x)$ along the interval (a,b)

$$\langle h(x),T_m(X_e)\rangle_e = \int_a^b \frac{2xh(x)T_m(X_e)dx}{\sqrt{(b^2-x^2)(x^2-a^2)}}$$
$$= \frac{\pi}{2}\frac{1}{\pi}\int_0^{2\pi} h\left(\sqrt{\frac{a^2+b^2}{2}+\frac{b^2-a^2}{2}\cos\theta}\right)$$
$$\times\cos(m\theta)d\theta$$

is obtained. Hence,

$$g_m^{(e)} = \left(\frac{\pi}{2}\right)^2 \hat{h}_m,$$

where $\hat{h}_m$ are the cosines Fourier coefficients of the even, 2π-periodic function

$$\hat{h}(\theta) = h\left(\sqrt{\frac{a^2+b^2}{2}+\frac{b^2-a^2}{2}\cos\theta}\right).$$

Also, for the coefficients

$$h_{mn}^{(e)} = \langle\langle h(x,x'),T_n(X_e')\rangle_e,T_m(X_e)\rangle_e$$
$$= \int_a^b\int_a^b \frac{4xx'T_m(X_e)T_n(X_e')h(x,x')dxdx'}{\sqrt{(b^2-x^2)(x^2-a^2)}\sqrt{(b^2-x'^2)(x'^2-a^2)}}$$

there results

$$h_{mn}^{(e)} = \frac{1}{4}\int_0^{2\pi}\int_0^{2\pi} h\left(\sqrt{\frac{a^2+b^2}{2}+\frac{b^2-a^2}{2}\cos\theta},\right.$$
$$\left.\sqrt{\frac{a^2+b^2}{2}+\frac{b^2-a^2}{2}\cos\theta}\right)\cdot\cos(m\theta)\cos(n\theta')d\theta d\theta',$$

a formula which can be used in the case an analytical expression is given for the function $h(x,x')$ by means of a 2D cosine Fourier transform.

The terms containing integrals are defined as

$$S_n^e(k_0t) = \langle\sin(k_0tx),T_n(X_e)\rangle_e$$
$$= \frac{1}{2}\int_0^{2\pi}\sin\left[k_0t\sqrt{\frac{a^2+b^2}{2}+\frac{b^2-a^2}{2}\cos\theta}\right]\cos(n\theta)d\theta.$$

Hence,

$$I_1^e \equiv \int_a^b\int_a^b dxdx'\int_0^2 \frac{4xx'\sin(k_0tx)T_m(X_e)}{\sqrt{(b^2-x^2)(x^2-a^2)}}$$
$$\times\frac{\sin(k_0tx')T_n(X_e')}{\sqrt{(b^2-x'^2)(x'^2-a^2)}}\frac{dt}{t^2-1}$$
$$= \int_0^2 \frac{S_n^e(k_0t)S_m^e(k_0t)}{t^2-1}dt = \sum_{j=1}^{2p}\frac{w_j}{t_j^2-1}S_n^e(k_0t_j)S_m^e(k_0t_j).$$

Finally

$$F_n = \frac{4}{\pi}\int_0^a dx\int_0^{2\pi} K^{(o)}$$
$$\times\left(x,\sqrt{\frac{a^2+b^2}{2}+\frac{b^2-a^2}{2}\cos\theta}\right)\cos(n\theta)d\theta$$

for the explicit part of the kernel and

$$IF_1 \equiv \frac{4}{\pi}\int_0^a dx\int_a^b \frac{2x'T_n(X_e')dx'}{\sqrt{(b^2-x'^2)(x'^2-a^2)}}$$
$$\times\int_0^2 \frac{\sin(k_0xt)\sin(k_0x't)}{t^2-1}dt = \frac{2}{\pi}$$
$$\times\int_0^2\left(\frac{1-\cos(k_0at)}{k_0t}\right)\frac{S_n^e(k_0t)}{t^2-1}dt$$

for the part containing integrals which cannot be obtained in closed form. The last integrals can be computed (for all the values of n) by using a vectorized form of an even Gauss-Legendre quadrature formula.

VII. NUMERICAL RESULTS

The parameters in the present problem are: the angles θ_i,φ_i giving the incident-wave direction, a, the diaphragm's half length, b-a the slit's width, the depth h, and frequency $f=\omega/2\pi$. We consider $a=1$ such that all the lengths are normalized with respect to half-diaphragm length.

For determining the pressure on the diaphragm we have to solve firstly the systems (34) and (39) for the coefficients $A_n^{(o)}$ and $A_n^{(e)}$. As the representation formulas (32) and (37) assure the proper behavior of the pressure at the points a and b the infinite systems have good convergence properties, such that only a few terms have to be retained in the infinite systems. Once the parameters $A_n^{(o)}$ and $A_n^{(e)}$ are determined, the moment M and the force F are provided by formulas (43) and (42), respectively.

The numerical values $M(\theta_i,\varphi_i)$ obtained by solving the problem satisfy the relationship

$$M(\theta_i,\varphi_i) = M\cos\varphi_i\sin\theta_i,$$

where

$$M = M(\pi/2,0).$$

We denote by Mo and Fo the moment and the force corresponding to the incoming and reflected wave when $b=a$ (the solution of the problem without slits) and have plotted in Fig. 4 the value of the ratio M/Mo for certain values of $h>0$ and $b>1$. Also, in Fig. 5 is plotted the amplitude of

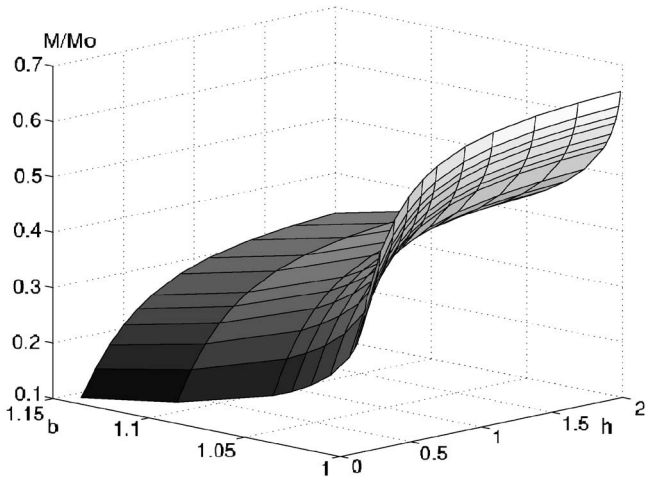


FIG. 4. Variation of the moment ratio M/M_o with b and h at a frequency of 10 kHz.

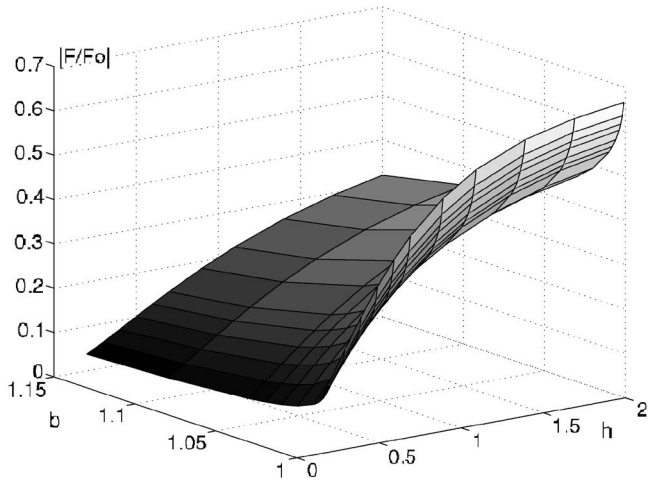


FIG. 5. Variation of the modulus of force ratio F/F_o with b and h at a frequency of 10 kHz.

the ratio F/Fo. It is to be noticed that in contrast to the moment, the force F has a phase delay plotted in Fig. 6.

Figures 4 and 5 point out that the interesting case from the design point of view, characterized by larger values of moment and force, is that of values of b closer to 1 and larger values of h. All the calculations were carried out for the frequency value $f=10$ kHz.

Some physical explanations of these results can be provided by analyzing the diffraction of a plane wave by a grating. This problem, in the case of a simplified one-mode approximation, has an explicit analytical solution[11,12] and a plot of the transmission coefficient is given in Ref. 13, Fig. 2(b). It is clear that the transmission coefficient is significantly lower than unity only in the case of very narrow slits. For wider slits, an important part of the incoming plane wave is passing through slits in the lower half-plane, equalizing the pressure on the two faces of the diaphragm. As the diaphragm is driven mainly by the pressure difference on the two faces of the diaphragm (or by the net moment due to the pressure difference) it is clear that the device will work better in the case of very narrow slits. In the case of the finite depth of the backing cavity, the lower wall of the backchamber will give a reflection of the waves which have penetrated the slits

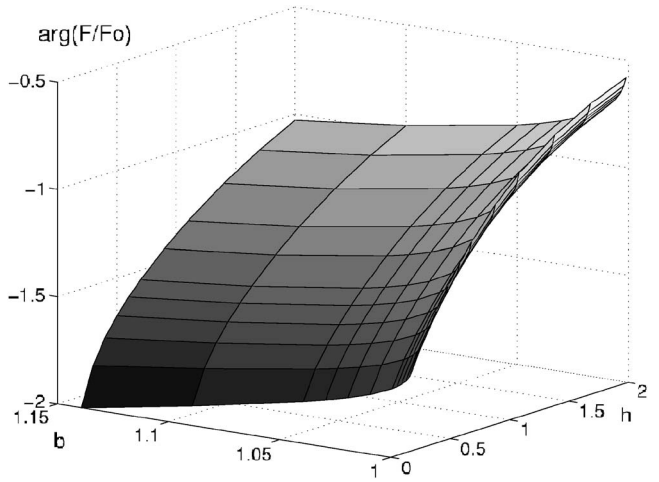


FIG. 6. Variation of the phase of force ratio F/F_o with b and h at a frequency of 10 kHz.

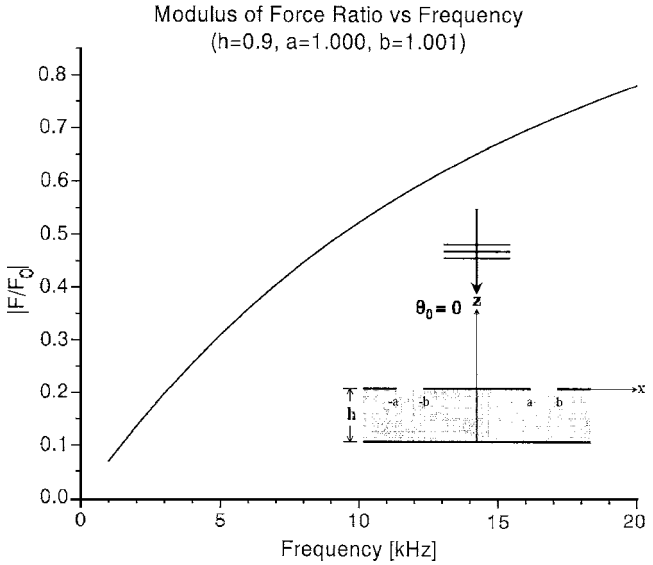


FIG. 7. Increase of the modulus of the force ratio with frequency ($a=1.0$ mm, $b=1.001$ mm, $h=0.9$ mm, $\theta_0=0$).

giving an additional increase of the pressure on the lower face of the diaphragm and decreasing the pressure difference between the two faces.

Next, we investigate the dependence of moment and force upon frequency. These functions are plotted in Figs. 7 and 8. The normalized force amplitude $|F/Fo|$ is increasing nearly linearly with frequency while the ratio M/Mo is practically independent of frequency.

The analysis performed in this paper considers a simplified model for the directional microphone. There are many aspects of the analysis and design to consider in the development of the device. This paper focusses only on the applied moments and forces due to sound. Finally, we assumed in this work that the diaphragm is a rigid plate. The elasticity

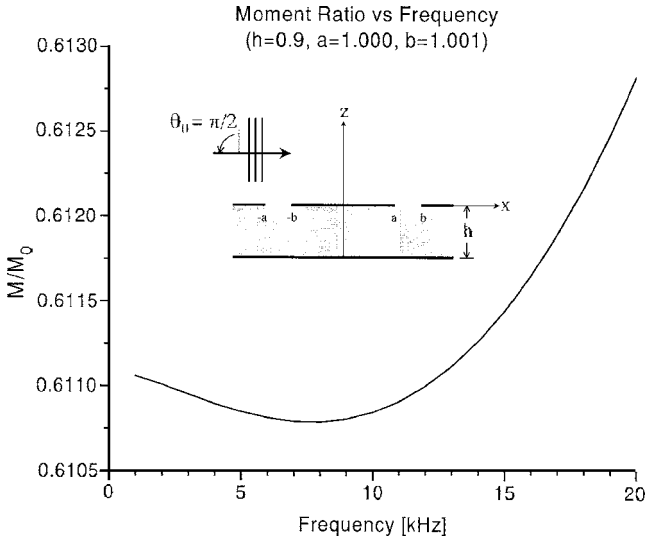


FIG. 8. Variation of the resultant moment of pressure with frequency ($a=1.0$ mm, $b=1.001$ mm, $h=0.9$ mm, $\theta_0=90°$).

of the real diaphragms will give supplementary problems which can be addressed only in a more completed computational model.

ACKNOWLEDGMENT

This work has been supported through NIH Grant No. R01 DC05762-1A1, and DARPA Grant No. DAAD17-00-C-0149 to R.N.M.

[1]R. N. Miles, R. Robert, and R. R. Hoy, "Mechanically coupled ears for directional hearing in the parasitoid fly Ormia Ochracea," J. Acoust. Soc. Am. **98**, 3059–3070 (1995).
[2]Q. Su, R. N. Miles, M. G. Weinstein, R. A. Miller, L. Tan, and W. Cui, "Response of a biologically inspired MEMS differential microphone diaphragm," in Proceedings of the SPIE AeroSense 2002, Orlando FL, paper No. [4743-15].
[3]R. N. Miles, S. Sundermurthy, C. Gibbons, D. Robert, and R. Hoy, Differential Microphone, United States Patent **6**, 788–796 (2004)
[4]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, Princeton, 1986).
[5]M. J. Ablowitz and A. S. Fokas, *Complex Variables: Introduction and Applications* (Cambridge University Press, Cambridge, 2000).
[6]F. P. Mechel, *Formulas of Acoustics* (Springer-Verlag, Berlin, 2002).
[7]M. Ochmann and F. P. Mechel, "Analytical and numerical methods in acoustics," in *Formulas of Acoustics*, edited by Mechel (Springer-Verlag, Berlin, 2002), Chap. O.
[8]I. S. Gradshtein and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 5th ed., English translation, edited by A. Jeffrey (Ed.). (Academic, New York, 1994).
[9]V. M. Aleksandrov, E. V. Kovalenko, and S. M. Mkhitarian, "Method of obtaining spectral relationships for integral operators of mixed problems of mechanics of continuous media," J. Appl. Math. Mech. **46**, 825–832 (1983).
[10]L. N. Trefethen, *Spectral Methods in Matlab* (SIAM, Philadelphia, 2000).
[11]J. W. Miles, "On Rayleigh scattering by a grating," Wave Motion **4**, 285–292 (1982).
[12]E. Scarpetta and M. A. Sumbatyan, "Explicit analytical results for one-mode oblique penetration into a periodic array of screens," IMA J. Appl. Math. **56**, 109–120 (1996).
[13]D. Homentcovschi, R. N. Miles, and L. Tan, "Influence of viscosity on the diffraction of sound by a periodic array of screens," J. Acoust. Soc. Am. **117**, 2761–2771 (2004).

Characterization of scatterer motion in a reverberant medium

Stéphane G. Conti[a)]
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California, 92037

Julien de Rosny
Laboratoire Ondes et Acoustique, 10 Rue Vauquelin, 75005, Paris, France

Philippe Roux
Marine Physical Laboratory, SIO/UCSD, 9500 Gilman Dr., La Jolla, California 92093-0205

David A. Demer
Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, La Jolla, California, 92037

(Received 20 April 2005; revised 9 November 2005; accepted 9 November 2005)

Recently, an acoustic technique has been proposed to measure the scattering strength and the dynamics of weak moving scatterers in a reverberant cavity: diffusing reverberant acoustic wave spectroscopy (DRAWS). Both parameters are obtained from the correlations of the reverberated-scattered transient pressure fields for different scatterers positions. This technique is based on a diffusive field theory [de Rosny *et al.*, Phys. Rev. Lett., **90**, 094302 (2003)]. Here, a more systematic approach of the DRAWS technique properties is presented. Moreover, an important extension is proposed using the fourth-order moment of the field, or the variance of the correlation estimator. Contrary to the correlations (second-order moment) that allow the measurement of the scattering cross section, its variance (fourth-order moment) is shown to be mainly sensitive to the scatterer displacement. The robustness of DRAWS is discussed using different configurations: a computer simulation of a moving scatterer in a two-dimensional cavity. Experiments were also carried out with a 23-mm-diameter copper sphere moved by stepping motors in 16 liters of water, and finally a human walking in a 125-m^3 reverberant room. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146107]

PACS number(s): 43.20.El, 43.20.Ye, 43.30.Gv, 43.80.Ev [EJS] Pages: 769–776

I. INTRODUCTION

In the last decades, techniques based on multiply scattered waves have been developed to characterize clouds of moving scatterers. Conventional techniques such as Doppler imaging cannot be applied in this case because they are based on a single scattering approximation. To overcome this limitation, the diffusing wave spectroscopy (DWS; Maret and Wolf, 1987; Pine *et al.*, 1988) has been developed in optics to measure the diffusion constant of Brownian scatterers from the temporal fluctuations of the speckle of a multiply scattered laser beam. Later, DWS was extended to acoustic waves with the diffusing acoustic wave spectroscopy (DAWS; Cowan *et al.*, 2000; 2002). In acoustics, one can easily record the transient acoustic response of the multiple scattering medium. Therefore, more information can be obtained from the field fluctuations with respect to the propagation length, i.e., the scattering order. Finally, for weak scatterers in motion in a strongly reverberant medium, it has been recently demonstrated that the acoustical total scattering cross section σ_T can be determined from an ensemble of transient pressure fields recorded for uncorrelated positions of the scatterer (de Rosny, 2000; de Rosny and Roux, 2001). It was shown that the ratio of the coherent intensity (square of the sum of the transient responses) and the incoherent intensity (sum of the squared transient responses) decreases as $\exp(-L/l_s)$, where L is the acoustic path, and l_s the mean free path. Later, a more general approach was developed in the case of a single scatterer (de Rosny *et al.*, 2003). Like DWS, it consists of studying the normalized correlations between transient fields recorded at a given propagation time for two different positions of one scatterer. These works led to the diffusing reverberant acoustic wave spectroscopy (DRAWS), where the dynamics and the total scattering cross section σ_T can be estimated from normalized field correlations. DRAWS is based on two main assumptions. First, the mean free path l_s of the scatterer is much larger than the dimensions of the cavity. Second, the wave is reflected multiple times on the boundary of the cavity before extinction. The last assumption implies that the field is diffuse in the sense of room acoustics theory (Schroeder, 1959), i.e., it is isotropic and homogeneous.

The accuracy and precision of DRAWS to measure the total scattering cross section σ_T was characterized using standard metal spheres in water for ultrasonic waves (Demer *et al.*, 2003). DRAWS was since used successfully to measure σ_T for Antarctic and Northern krill in small seawater containers with ultrasonic waves (Demer and Conti, 2003; Conti *et al.*, 2005), for fish in large seawater tanks with ultrasonic waves (Conti and Demer, 2003), and for humans in air with audible acoustic waves (Conti *et al.*, 2004).

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

However, the ability of DRAWS to characterize the average scatterer motion was investigated only once by de Rosny *et al.* (2003). Here, we propose a more detailed investigation of DRAWS to characterize the type and magnitude of the scatterer displacement. Also, we extend the DRAWS theory and demonstrate that the variance of the correlated field for two positions of a scatterer can provide a metric for the magnitude of the scatterer displacement, independently of the total scattering cross section.

The theoretical principles are confirmed first using simulations for a single scatterer in a two-dimensional reverberant cavity, and then experimentally with a standard metal sphere in a stainless-steel bucket filled with water for ultrasonic waves, and with a human in a squash court for the audible range.

II. FORMALISM

A. Theory

Consider a reverberant cavity of volume V with one fixed transducer transmitting N pulses at a constant repetition rate. For each pulse, transient fields in the cavity are recorded by one or more receivers. One scatterer is moving by constant steps δx in the cavity between each of the N pulses. The average displacement of the scatterer after n steps is denoted as $\delta r(n)$. For the pulse k, the transient field is denoted as $h_k(t)$, with k ranging from 1 to N. After n steps, the coherent intensity (the cross correlation) and the incoherent intensity (the autocorrelation) of the $h_k(t)$ are, respectively, $I_n(t) = \langle h_k(t) h_{k+n}(t) \rangle$ and $A(t) = \langle h_k(t)^2 \rangle$, where $\langle \cdot \rangle$ is the average over the N pulses (Ishimaru, 1978; Sheng, 1995).

When the positions of the scatterer are uncorrelated between transient fields, de Rosny and Roux (2001) demonstrated that the normalized cross-correlation function $g_1^t(n=1) = [I_1(t)]/[A(t)]$ of $h_k(t)$ and $h_{k+1}(t)$, decreases exponentially with the scattering mean free path l_s: $g_1^t(n=1) \approx \exp(-tc/l_s)$, where c is the sound speed in the medium. The time decay of $g_1^t(n)$ is a mean to estimate the scattering mean free path l_s, and the total scattering cross section of the scatterer for a diluted medium since $\sigma_T = V/l_s$.

It is shown here that the normalized cross-correlation function $g_1^t(n) = [I_n(t)]/[A(t)]$ is a function of both the total scattering cross section σ_T and the displacement $\delta r(n)$ of the scatterer. Then, we demonstrate that the variance $\Sigma_{In}(t) = \langle (h_k(t) h_{k+n}(t))^2 \rangle - \langle h_k(t) h_{k+n}(t) \rangle^2$ of the cross correlation $h_k(t) h_{k+n}(t)$ can be written as a function of $g_1^t(n)$ and the total scattering cross section σ_T, and used to estimate the average displacement of the scatterer.

1. Normalized cross-correlation function

The expression of the normalized cross-correlation function with the motion of the scatterer was presented by de Rosny *et al.* (2003). The acoustic field $h_k(t)$ can be described as the infinite summation of the contributions $h_k^m(t)$, corresponding to the field scattered m times. In a dilute medium, the contributions $h_k^m(t)$ are uncorrelated between each other, since the scattering events are independent,

$$h_k(t) = \sum_{m=0}^{\infty} h_k^m(t).$$

The total intensity after m scattering events is proportional to $\langle h_k^m(t)^2 \rangle$. Between t and $t+dt$, the total intensity in the cavity will decrease by $(c/l_s)\langle h_k^m(t)^2 \rangle dt$ due to the scattering from the scatterer after m scattering events, and also by $(c/l_a) \times \langle h_k^m(t)^2 \rangle dt$ due to the absorption from the cavity and the scatterer. l_a is the absorption mean free path due to the attenuation in the cavity and the absorption from the scatterer. At the same time, after the first scattering event has occurred $m > 0$, the total intensity increases by $(c/l_s)\langle h_k^{m-1}(t)^2 \rangle dt$ corresponding to the intensity from the $m-1$ preceding scattering events. From these considerations, the following Kolmogoroff system of equations can be written:

$$\frac{d\langle h_k^0(t)^2 \rangle}{dt} = -\left(\frac{c}{l_s} + \frac{c}{l_a} \right) \langle h_k^0(t)^2 \rangle \tag{1}$$

and for $m > 0$

$$\frac{d\langle h_k^m(t)^2 \rangle}{dt} = \frac{c}{l_s} (\langle h_k^{m-1}(t)^2 \rangle - \langle h_k^m(t)^2 \rangle) - \frac{c}{l_a} \langle h_k^m(t)^2 \rangle. \tag{2}$$

The solution for this system is a Markoff–Poisson process

$$\langle h_k^m(t)^2 \rangle = \langle h_k^0(t)^2 \rangle \exp\left(-\frac{tc}{l_s} \right) \frac{(tc/l_s)^m}{m!} \exp\left(-\frac{tc}{l_a} \right). \tag{3}$$

From the independence of the scattering events, $\langle h_k^m(t) h_k^{m'}(t) \rangle = 0$ for $m \neq m'$, the coherent intensity becomes

$$I_n(t) = \langle h_k(t) h_{k+n}(t) \rangle = \sum_{m=0}^{\infty} \langle h_k^m(t) h_{k+n}^m(t) \rangle. \tag{4}$$

The normalized cross-correlation function of the mth scattering event is

$$g_1^{(m)}(n) = \frac{\langle h_k^m(t) h_{k+n}^m(t) \rangle}{\langle h_k^m(t)^2 \rangle}; \tag{5}$$

therefore,

$$\langle h_k(t) h_{k+n}(t) \rangle = \sum_{m=0}^{\infty} \langle h_k^m(t) h_{k+n}^m(t) \rangle = \sum_{m=0}^{\infty} \langle h_k^m(t)^2 \rangle g_1^{(m)}(n), \tag{6}$$

and with Eq. (3)

$$\langle h_k(t) h_{k+n}(t) \rangle = \langle h_k^0(t)^2 \rangle \exp\left(-\frac{tc}{l_s} - \frac{tc}{l_a} \right) \times \left(1 + \sum_{m=1}^{\infty} \frac{(tc/l_s)^m}{m!} g_1^{(m)}(n) \right), \tag{7}$$

with $\langle h_k(t)^2 \rangle = \langle h_k^0(t)^2 \rangle \exp(-tc/l_a)$,

$$g_1^t(n) = \exp\left(-\frac{tc}{l_s} \right) \left(1 + \sum_{m=1}^{\infty} \frac{(tc/l_s)^m}{m!} g_1^{(m)}(n) \right). \tag{8}$$

The normalized cross-correlation $g_1^{(m)}(n)$ is equal to the integral over the possible motion y of the normalized spatial

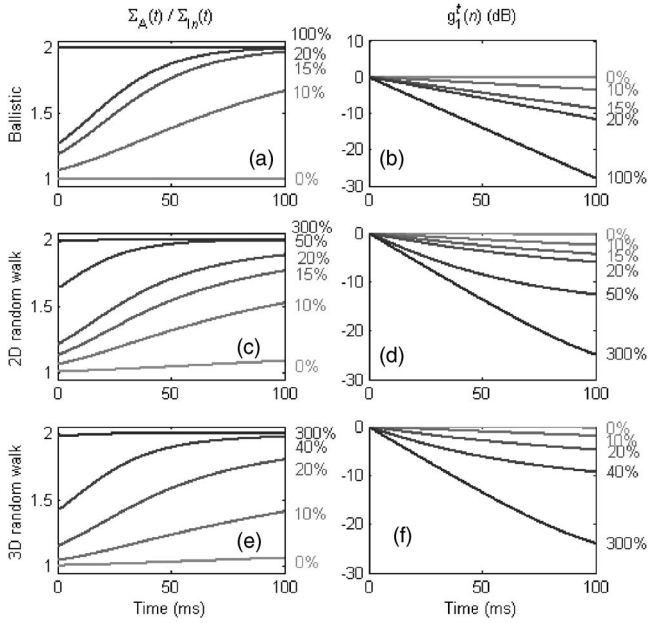


FIG. 1. Theoretical predictions for the ratio $\Sigma_A(t)/\Sigma_{In}(t)$ and $g_1^t(n)$ in the case of a 23-mm copper sphere in 16 liters of water and various displacement amplitudes. (a) and (b): Case of ballistic motion. (c) and (d): Case of 2D random walk. (e) and (f): Case of 3D random walk. The displacement of the scatterer relative to wavelength $\delta x/\lambda$ is comprised between 0% and 300%, from light to dark.

cross-correlation function of the wave field $S_c(y)$ multiplied by the probability for the scatterer to move a distance y after n steps,

$$g_1^{(m)}(n) = \int_0^\infty S_c(y)^m P_n(y) dy. \tag{9}$$

For ballistic motion $\delta r(n) = n\delta x$, and the normalized spatial cross-correlation function of the wave field is $S_c(y) = [\sin(2\pi y/\lambda)/(2\pi y/\lambda)]^2$ (Cook *et al.*, 1955), and the probability $P_n(y)$ is the Dirac function $\delta(y - \delta r(n))$ (de Rosny *et al.*, 2003). In this case, the normalized cross correlation is given by

$$g_1^{(m)}(n) = \exp\left(-\frac{m(2\pi n\delta x/\lambda)^2}{3}\right). \tag{10}$$

Finally, for ballistic motion, the normalized cross-correlation function $g_1^t(n)$ becomes

$$g_1^t(n) = \exp\left(-\frac{tc}{l_s}\left(1 - \exp\left(-\frac{1}{3}\left(2\pi\frac{n\delta x}{\lambda}\right)^2\right)\right)\right). \tag{11}$$

It can be rewritten as a decreasing exponential: $g_1^t(n) = \exp(-\alpha(n)t)$, where the exponential decay with time of $g_1^t(n)$ for ballistic motion [Fig. 1(b)] is given by

$$\alpha(n) = \frac{c\sigma_T}{V}\left(1 - \exp\left(-\frac{1}{3}\left(2\pi\frac{n\delta x}{\lambda}\right)^2\right)\right). \tag{12}$$

In the case of a 2D or 3D random walk with steps δx, the normalized cross-correlation $g_1^{(m)}(n)$ becomes (de Rosny *et al.*, 2003)

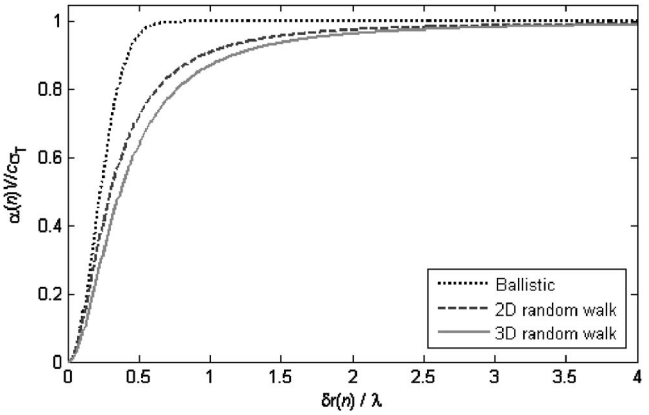


FIG. 2. Theoretical predictions of $\alpha(n)V/c$ normalized to the absolute value σ_T versus displacement of the scatterer $\delta r(n)$ relative to the wavelength λ. Case of ballistic motion (dark dashed) from Eq. (11), 2D random walk (light dashed) from Eq. (16), and 3D random walk (light solid) from Eq. (17).

$$g_1^{(m)}(n) = \frac{1}{1 + \frac{1}{3}m\left(2\pi\frac{\sqrt{n}\delta x}{\lambda}\right)^2} \tag{13}$$

and

$$g_1^{(m)}(n) = \frac{1}{1 + \frac{2}{9}m\left(2\pi\frac{\sqrt{n}\delta x}{\lambda}\right)^2}, \tag{14}$$

respectively. In these cases, no analytical expression can be derived for $g_1^t(n)$. Nevertheless, numerical estimates can be obtained from Eq. (8) [Figs. 1(d) and 1(f)]. The decay with time of $g_1^t(n)$ is not simply exponential anymore. However, for short propagation times, the slope of the semilogarithmic curve is fairly linear. A first-order approximation with tc/l_s yields the exponential decay $\alpha(n)$. Indeed,

$$g_1^t(n) \approx 1 - \frac{tc}{l_s}(1 - g_1^{(1)}(n)) \approx 1 - \alpha(n)t. \tag{15}$$

For a 2D or 3D random walk, $\alpha(n)$ becomes

$$\alpha(n) = \frac{c\sigma_T}{V}\left(1 - \left[1 + \frac{1}{3}(2\pi\sqrt{n}\delta x/\lambda)^2\right]^{-1}\right), \tag{16}$$

and

$$\alpha(n) = \frac{c\sigma_T}{V}\left(1 - \left[1 + \frac{2}{9}(2\pi\sqrt{n}\delta x/\lambda)^2\right]^{-1}\right), \tag{17}$$

respectively. Therefore, measuring the slope at the origin with respect to n is a very convenient and simple way to characterize the motion of the scatterer. For each of the three types of motion, the theoretical predictions for the measurement $\alpha(n)V/c$ are shown in Fig. 2. $\alpha(n)V/c$ tends asymptotically to the absolute σ_T, as the average displacement between acquisitions increases. Hence, the estimate of the total cross section using $\alpha(n)V/c$ is within 4% of the expected value σ_T when the displacement of the scatterer is greater than 0.5, 2, or 3 times the wavelength respectively for ballistic motion, a 2D random walk, or a 3D random walk (Table I).

TABLE I. Percentage of the measurement $\alpha V/c\sigma_T$ for various threshold values of the scatterer displacement δr relative to the wavelength λ. Case of ballistic motion and 2D or 3D random walks.

Motion type	Ballistic	Random walk 2D	Random walk 3D
$\delta r/\lambda$	0.5	2	3
$\alpha V/c\sigma_T$	96.5%	96.7%	97.7%

2. Ratio of the variances

It is of interest to study $\Sigma_{In}(t)$, the variance of the cross correlation $h_k(t)h_{k+n}(t)$. Indeed, we demonstrate that the ratio of the variances $\Sigma_{I0}(t)/\Sigma_{In}(t)$ is mainly sensitive to the scatterer displacement, and not to the total scattering cross section. The variance estimator $\Sigma_{In}(t)$ is given by

$$\Sigma_{In}(t) = \frac{1}{N-n}\sum_{k=1}^{N-n} h_k(t)^2 h_{k+n}(t)^2 - \left(\frac{1}{N-n}\sum_{k=1}^{N-n} h_k(t)h_{k+n}(t)\right)^2. \quad (18)$$

Assuming the $h_k(t)$ are jointly Gaussian variables, and introducing the variance of the autocorrelation $\Sigma_A(t)$ which is equal to $\Sigma_{I0}(t)$, the ratio of the variances is (see the Appendix)

$$\frac{\Sigma_A(t)}{\Sigma_{In}(t)} = 2\frac{1-\exp\left(-2\frac{tc}{l_s}\right)}{1+g_1^t(n)^2-2\exp\left(-2\frac{tc}{l_s}\right)}. \quad (19)$$

For ballistic motion, $g_1^t(n)$ with Eq. (11) gives the following analytical expression for the ratio of the variances:

$$\frac{\Sigma_A(t)}{\Sigma_{In}(t)} = 2\frac{1-\exp\left(-2\frac{tc}{l_s}\right)}{1-\exp\left(-2\frac{tc}{l_s}\right)\left[2-\exp\left(2\frac{tc}{l_s}\exp\left(-\frac{1}{3}\left(2\pi\frac{n\delta x}{\lambda}\right)^2\right)\right)\right]}, \quad (20)$$

as shown in Fig. 1(a).

For 2D or 3D random walks, the analytical formulas for the ratio $\Sigma_A(t)/\Sigma_{In}(t)$ cannot be obtained directly. But, $\Sigma_A(t)/\Sigma_{In}(t)$ can be evaluated numerically, as shown in Figs. 1(c) and 1(e) using Eqs. (13) and (14), respectively.

When the displacement of the scatterer is small, $g_1^t(n)=1$, and $\Sigma_A(t)/\Sigma_{In}(t)$ is slightly larger than 1. When the positions of the scatterer are uncorrelated, $g_1^t(n)=\exp(-tc/l_s)$, and the ratio $\Sigma_A(t)/\Sigma_{In}(t)$ reaches 2.

Since $g_1^t(n)$ depends mainly on the total scattering cross section, its ability to detect a weak scatterer will depend on the total scattering cross section if the displacement of the scatterer is large relative to the wavelength. On the other end, $\Sigma_A(t)/\Sigma_{In}(t)$ mainly depends on the displacement of the scatterer and not on the total scattering cross section. Therefore, this ratio is much more sensitive to a weak scatterer for large displacements. For instance, with either a 23-mm-diameter

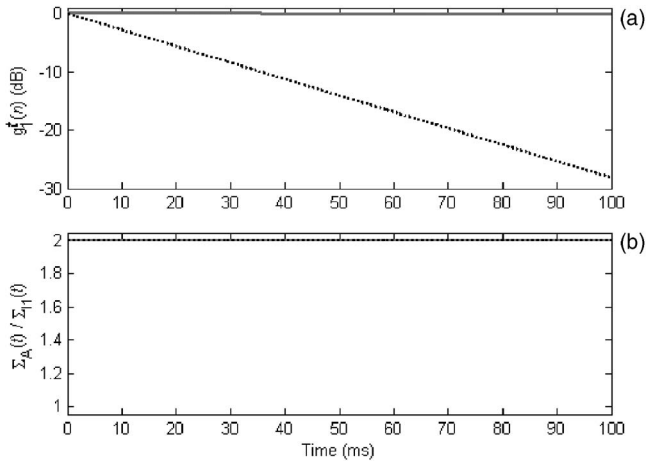


FIG. 3. Theoretical predictions for (a) $g_1^t(n=1)$ and (b) $\Sigma_A(t)/\Sigma_{In}(t)$ for a 23-mm-diameter copper sphere (σ_T=766 mm^2, dark dashed) and a 2-mm-diameter copper sphere (σ_T=5.78 mm^2, light solid) in 16 liters of water. Case of ballistic motion with a displacement of the scatterer relative to wavelength $\delta x/\lambda=3$.

(σ_T=766 mm^2) or a 2-mm-diameter copper sphere (σ_T =5.78 mm^2) in 16 liters of water, following ballistic motion with steps equal to 3 wavelengths, the theoretical predictions for the ratio $\Sigma_A(t)/\Sigma_{In}(t)$ are equal to 2, independently of the total scattering cross section, whereas the decay of $g_1^t(n)$ is too small to be estimated for the small sphere (Fig. 3). Hence, $\Sigma_A(t)/\Sigma_{In}(t)$ is a good metric to detect the motion of a weak scatterer. From a practical point of view, the use of either $g_1^t(n)$ or $\Sigma_A(t)/\Sigma_{In}(t)$ presents different advantages. $\Sigma_A(t)/\Sigma_{In}(t)$ is more sensitive to noise since it is defined from the fourth-order moments. The signal-to-noise ratio is the dominant parameter for practical use of the ratio $\Sigma_A(t)/\Sigma_{In}(t)$. In any case, a sufficiently large signal-to-noise ratio must also be obtained in order to estimate σ_T accurately using $g_1^t(n)$. Therefore, the condition on the signal-to-noise ratio must be verified for both $g_1^t(n)$ and $\Sigma_A(t)/\Sigma_{In}(t)$.

B. Displacement estimation

As shown by Eq. (8), the normalized cross-correlation function $g_1^t(n)$ depends on both the scattering mean free path l_s and the ratio between the displacement of the scatterer δx and the acoustic wavelength λ. If $\delta x/\lambda \gg 1$, the dependence of $g_1^t(n)$ on displacement of the scatterer is negligible. Therefore, when the displacement of the scatterer is large compared to the wavelength, the successive positions are uncorrelated, and the exponential decay of $g_1^t(n)$ leads to the total scattering cross section $\sigma_T = \alpha(n)V/c$ [see Eqs. (12), (16), and (17); Fig. 2]. For smaller displacements of the scatterer, $\delta x/\lambda$ lower than 1, the estimate of the exponential decay $g_1^t(n)$ is a function of n and $\delta x/\lambda$.

As discussed in part Sec. II A, the displacement of the scatterer can be estimated from the measurements of $\alpha(n)$. After n steps of amplitude δx, the actual displacement to consider for $g_1^t(n)$ is $n\delta x$ for ballistic motion, and $2\sqrt{n}\delta x$ or $\sqrt{6n}\delta x$ for 2D or 3D random walks, respectively. A minimization function $D(\delta x)$ can be defined in order to estimate the displacement from these measurements,

$$D(\delta x) = \frac{V}{c\sigma_T} \sum_{n=1}^{n_{\max}} |\bar{\alpha}(n) - \alpha(n, \delta x)|, \tag{21}$$

where $\bar{\alpha}(n)$ is the value measured from the data.

This generalization of the DRAWS method in regard to the fourth-order moments and the estimate of the displacement of the scatterer is tested and confirmed here using experiments for which the displacements of the scatterer are controlled. First, a standard metal sphere is moved by stepping motors. Then, a human walks with controlled steps in a reverberant room. In the case of a group of scatterers, in the limit of a large number of scatterers compared to the scattering order m, $g_1^{(m)}(n)$ is equal to $(\int_0^\infty S_c(y)P_n(y)dy)^m$ (de Rosny *et al.*, 2003).

III. RESULTS

A. Simulations

The simulation was performed using a second-order finite element time-domain scheme for a two-dimensional cavity, similar to the one presented in de Rosny and Roux (2001). There was no absorption in the cavity, therefore $l_a = \infty$. The cavity was a 300- by 300-pixel square, with a spatial resolution of 8 pixels per wavelength (0.125λ per pixel), and Dirichlet boundaries conditions. The sound speed was $c = 1000$ m/s. One emitter transmitted a 50% bandwidth pulse with center frequency $f_c = 1$ kHz. Reverberation time series were recorded using 28 receivers, 2λ apart from each other. The scatterer in the cavity corresponded to 1 pixel, and was defined by the sound-speed contrast $c_{\text{scatterer}}/c$. Since the total scattering cross section of the scatterer σ_T could not be derived theoretically for the simulation due to the heterogeneity of the two-dimensional simulation lattice, σ_T was evaluated using simulations measuring l_s for $K = 0$ to 50 scatterers in the cavity taking random uncorrelated positions. σ_T was obtained from the slope of V/l_s versus number of scatterers since $\sigma_T(K) = V/Kl_s$, with $V = (300 \times 0.125)^2$. For a sound-speed contrast $c_{\text{scatterer}}/c = 0.7$, the total scattering cross section normalized to the wavelength σ_T/λ was 3.122×10^{-2}.

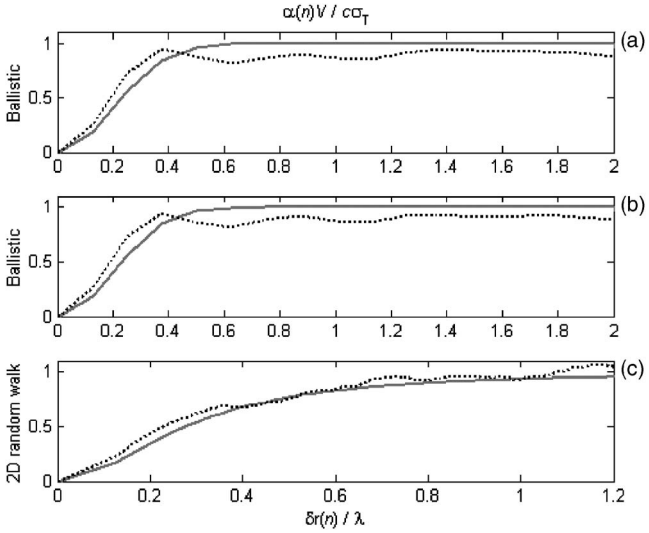


FIG. 4. Measurement of $\alpha(n)V/c$ versus displacement relative to wavelength $\delta r(n)/\lambda$. Case of a 2D simulation for a single scatterer with sound-speed contrast $c_{\text{scatterer}}/c = 0.7$. The measurements (dark dashed) are compared to theoretical predictions (light solid) for (a) and (b) two different ballistic motions, and (c) a 2D random walk. The theoretical predictions are obtained using Eq. (11) for the ballistic motions and Eq. (16) for the random walk.

The same simulation was then used with only one scatterer for two different ballistic displacements and a random walk. In all three cases, the scatterer was moved 1 pixel between consecutive pulses k and $k+1$, corresponding to a displacement $\delta x = \lambda/8$. Ensembles of 200 positions were simulated, but $N = 100$ positions was used for the calculations. Using the pulses k and $k+n$ as consecutive, with n ranging from 1 to 100, the displacement of the scatterer was $\delta r(n) = n\lambda/8$ for the ballistic motion. For the random walk, the average displacement of the scatterer was estimated using the actual positions of the scatterer during the simulation.

The theoretical predictions and the measurements of $\alpha(n)$ versus relative displacement of the scatterer for the two ballistic motions, and the 2D random walk are in good agreement (Fig. 4).

B. Controlled experiments

1. Sphere moved by stepping motors

A first set of experiments was obtained using a 23-mm-diameter copper sphere in a 20-liter stainless-steel bucket filled with approximately 16 liters of fresh water. The transmitted signal was a 15-ms-long chirp between 200 and 800 kHz ($f_c = 500$ kHz, $\lambda/2 = 1.4$ mm), and the transient fields were recorded over 50 ms using an 8-bit resolution oscilloscope for five positions of the receiver. The total scattering cross section was estimated for three narrow frequency bands centered at 300, 500, and 700 kHz with a 10-kHz bandwidth, after filtering the data in the corresponding narrow bands. The sphere was moved using stepping motors. The displacement of the sphere between two consecutive transient fields was $\delta x = 0.1$ mm for a total of 200 positions

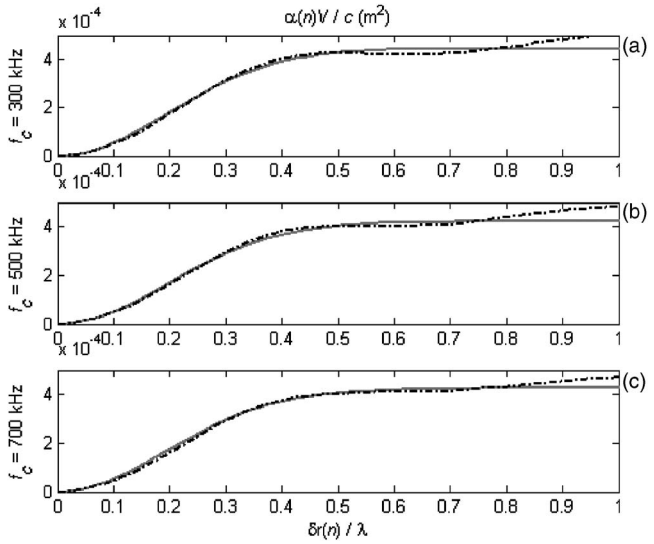


FIG. 5. Estimate of $\alpha(n)V/c$ versus displacement relative to wavelength $\delta r(n)/\lambda$ for a 23-mm-diameter copper sphere following ballistic motion in 16 liters of water. The measurements (dark dashed) are obtained for 10-kHz frequency bands centered at (a) 300; (b) 500; and (c) 700 kHz. Theoretical predictions (light solid) are obtained using Eq. (11).

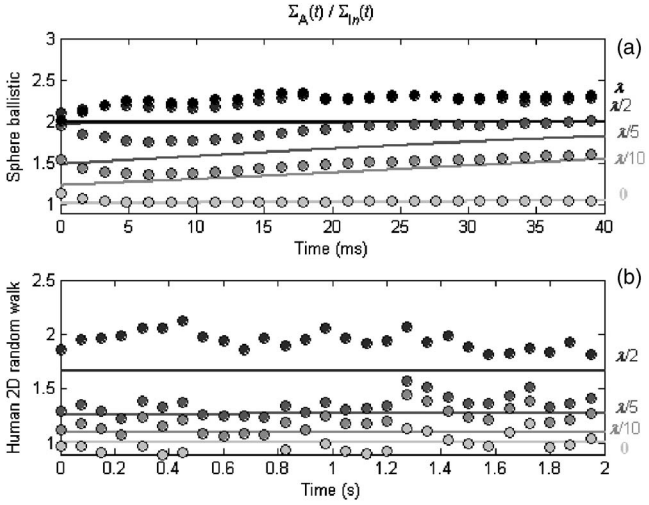


FIG. 6. Ratio of the variances $\Sigma_A(t)/\Sigma_{In}(t)$, experimental results for (a) a 23-mm-diameter copper sphere following a ballistic motion in 16 liters of water at 700 kHz; (b) a human in a squash court for a 2D random walk with 3-cm steps. The points correspond to the experimental measurements, and the straight lines to the theoretical predictions from Eq. (20) for the ballistic motion, and Eqs. (19) and (13) for random walk. The displacement of the scatterer relative to wavelength $\delta x/\lambda$ is 0; 1/10; 1/5; 1/2; and 1, from light to dark.

($\delta x < \lambda/20$). Practically, the sphere was moved constantly at a very low speed (0.02 mm/s), and its position was measured continuously with the stepping motor controller. Once the desired position was reached, the emission/acquisition system was triggered, and the measured position of the sphere was compared to the desired one.

The results are in good agreement with the theory for each of the frequency bands centered at 300, 500, and 700 kHz (Fig. 5). For displacement greater than half the wavelength, the measured total scattering cross section corresponds to the expected value. For $\delta r/\lambda > 0.9$ (approximately), the measurements appear to be higher than the expected value. In order to achieve such displacements of the sphere, the time intervals between shots had to be increased. As the time intervals increase, parameters of the medium such as the temperature are more likely to fluctuate. These fluctuations lead to sound-speed fluctuations, and a positive bias in the measurements explaining the higher values for large displacements of the scatterer.

The experimental results for the ratio of the variances $\Sigma_A(t)/\Sigma_{In}(t)$ with the 23-mm copper sphere following ballistic motion at 700 kHz shows a good agreement with theory [Fig. 6(a)]. As the time interval between the time series was increased for the large displacements of the sphere, some discrepancies appeared between the theoretical predictions and the experimental results.

2. Human walking in a reverberant room

A second set of experiments was realized with a human walking in a squash court using the experimental setup presented in Conti *et al.* (2004). Two-second chirps were transmitted between 100 and 500 Hz (f_c=300 Hz, $\lambda/2$=57.5 cm), and ensembles of 100 transient fields were recorded over 5 s using four receivers. During the experiment, a human walked in the room either along a straight line, or following a random walk. In both cases, the human moved between pulses by small steps of δx=3 cm ($\delta x < \lambda/20$), with a 0.5-cm precision (approximately).

For both ballistic motion and 2D random walk, the experimental results with the human walking in the squash court are in good agreement with theory (Fig. 7). For the 2D random walk, the number of transient fields recorded did not provide a larger range for the displacement of the scatterer than the one presented. But, even over this limited range, the measurements and the theory are in good agreement, and sufficient to estimate the average displacement of the human.

For the ratio of the variances $\Sigma_A(t)/\Sigma_{In}(t)$ with the hu-

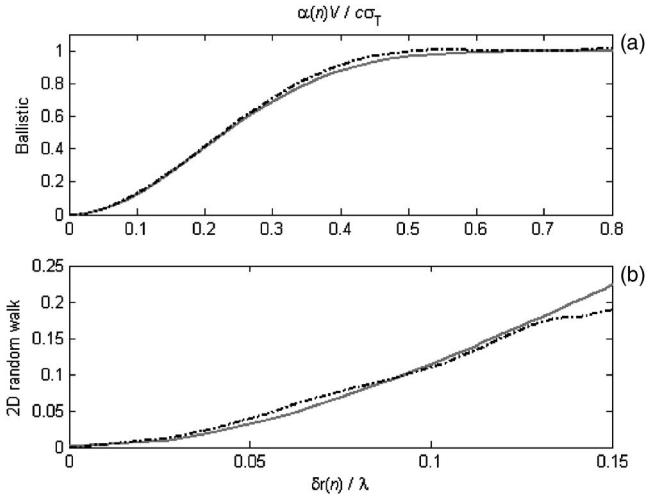


FIG. 7. Normalized measurement of $\alpha(n)V/c\sigma_T$ versus displacement relative to wavelength $\delta r(n)/\lambda$ for a human in a squash court with 3-cm steps. (a) Case of ballistic motion. (b) Case of 2D random walk. The measurements (dark dashed) are obtained between 100 and 500 Hz. Theoretical predictions (light solid) are obtained using Eq. (11) for the ballistic motion, and Eq. (16) for the 2D random walk.

TABLE II. Estimated and expected displacement between shots obtained using the minimization of the function $D(\delta x)$.

Experiment	Estimated displacement	Expected displacement
Copper sphere 300 kHz	0.102 mm	0.1 mm
Copper sphere 500 kHz	0.100 mm	0.1 mm
Copper sphere 700 kHz	0.098 mm	0.1 mm
Human ballistic motion	3.26 cm	3 cm
Human 2D random walk	3.14 cm	3 cm

man following a 2D random walk [Fig. 6(b)], the results are similar to the ones obtained for the 23-mm-diameter copper sphere.

3. Estimating the displacement of the scatterer

Using the function $D(\delta x)$ from Eq. (21), the displacement of the 23-mm copper sphere and the human can be estimated precisely from the acoustical measurements (Table II). With the sphere, $D(\delta x)$ reaches a minimum within a few percent of the actual displacement used during the experiments, independent of the frequency (Fig. 8). Similar results are achieved for the experiments in the squash court with the human (Fig. 9). For the ballistic motion, the estimated displacement is less than 10% higher than the expected value, but the precision on the positions of the human was greater than 10% of the actual displacement (Table II). For the 2D random walk, the estimated displacement corresponds to the expected one within 5%.

IV. CONCLUSION

Using both simulations and experiments in different environments, the dependence with motion of the normalized cross-correlation function $g_1^t(n)$ in a reverberant medium has been demonstrated. These results are obtained for a scatterer following either ballistic motion or random walk. The correlation of the scatterer positions between time series can be evaluated using the ratio of the variances $\Sigma_A(t)/\Sigma_{In}(t)$.

Precise estimates of the displacement are obtained with a metal sphere for a controlled ballistic motion, and with a human for both ballistic motion and 2D random walk. The generalized DRAWS method could be used to monitor the activity of fish in an aquaculture facility remotely.

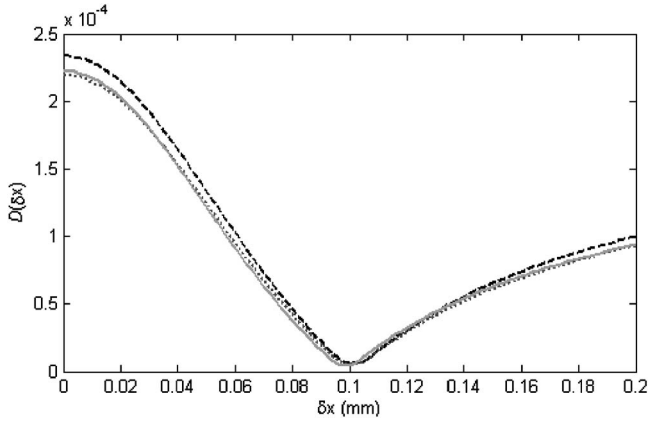


FIG. 8. Function $D(\delta x)$ to estimate the average displacement δx of the 23-mm copper sphere following ballistic motion in 16 liters of water. The results are obtained for three frequency bands centered at 300 (dark dashed), 500 (light dashed), and 700 (light solid) kHz, with a 10-kHz bandwidth.

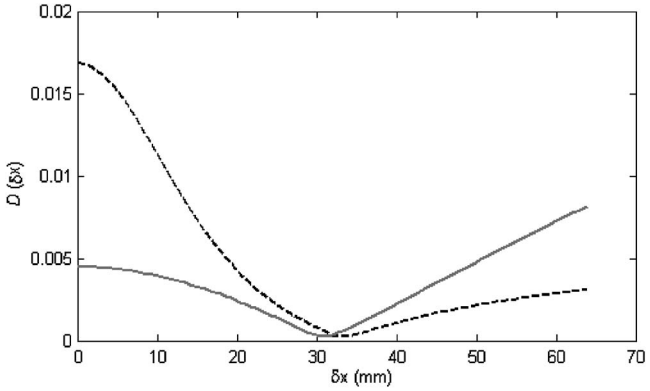


FIG. 9. Function $D(\delta x)$ to estimate the average displacement δx for the human walking in a squash court for ballistic motion (dark dashed), and 2D random walk (light solid).

APPENDIX: THEORETICAL DERIVATION OF THE VARIANCE ESTIMATOR

The variance estimator $\Sigma_{In}(t)$ of the cross correlation $h_k(t)h_{k+n}(t)$ is given by

$$\Sigma_{In}(t) = \frac{1}{N-n}\sum_{k=1}^{N-n} h_k(t)^2 h_{k+n}(t)^2 - \left(\frac{1}{N-n}\sum_{k=1}^{N-n} h_k(t)h_{k+n}(t)\right)^2. \quad \text{(A1)}$$

It can be written as

$$\Sigma_{In}(t) = \langle h_k(t)^2 h_{k+n}(t)^2\rangle - \frac{1}{(N-n)^2}\sum_{k,l=1}^{N-n} h_k(t)h_{k+n}(t)h_l(t)h_{l+n}(t) = \langle h_k(t)^2 h_{k+n}(t)^2\rangle - \langle h_k(t)h_{k+n}(t)h_l(t)h_{l+n}(t)\rangle. \quad \text{(A2)}$$

The moment theorem can be applied to the Gaussian random variables $h_k(t)$, $h_{k+n}(t)$, $h_l(t)$, and $h_{l+n}(t)$, and

$$\langle h_k(t)h_{k+n}(t)h_l(t)h_{l+n}(t)\rangle = \langle h_k(t)h_{k+n}(t)\rangle\langle h_l(t)h_{l+n}(t)\rangle + \langle h_k(t)h_l(t)\rangle\langle h_{k+n}(t)h_{l+n}(t)\rangle + \langle h_k(t)h_{l+n}(t)\rangle\langle h_{k+n}(t)h_l(t)\rangle. \quad \text{(A3)}$$

Each term of this sum can be expressed using the normalized cross-correlation function $g_1^t(n)$ as follows:

$$\langle h_k(t)h_{k+n}(t)\rangle\langle h_l(t)h_{l+n}(t)\rangle = \langle h_k(t)^2\rangle^2 g_1^t(n)^2, \quad \text{(A4)}$$

$$\langle h_k(t)h_l(t)\rangle\langle h_{k+n}(t)h_{l+n}(t)\rangle = \langle h_k(t)^2\rangle^2 g_1^t(\infty)^2, \quad \text{(A5)}$$

$$\langle h_k(t)h_{l+n}(t)\rangle\langle h_l(t)h_{k+n}(t)\rangle = \langle h_k(t)^2\rangle^2 g_1^t(\infty)^2. \quad \text{(A6)}$$

With $k=l$ in Eq. (24)

$$\langle h_k(t)^2 h_{k+n}(t)^2\rangle = \langle h_k(t)^2\rangle^2 + 2\langle h_k(t)h_{k+n}(t)\rangle^2$$
$$= \langle h_k(t)^2\rangle^2(1 + 2g_1^t(n)^2). \quad \text{(A7)}$$

The variance estimator $\Sigma_{In}(t)$ can be written as

$$\Sigma_{In}(t) = \langle h_k(t)^2\rangle^2(1 + g_1^t(n)^2 - 2g_1^t(\infty)^2). \quad \text{(A8)}$$

Since $g_1^t(\infty)=\exp(-tc/l_s)$ and $g_1^t(0)=1$ from Eq. (8), the variance estimator $\Sigma_{In}(t)$ becomes

$$\Sigma_{In}(t) = \langle h_k(t)^2\rangle^2\left(1 + g_1^t(n)^2 - 2\exp\left(-2\frac{tc}{l_s}\right)\right). \quad \text{(A9)}$$

Finally, introducing $\Sigma_A(t)=\Sigma_{I0}(t)$, the variance of the autocorrelation $h_k(t)^2$, the ratio of the variances is

$$\frac{\Sigma_A(t)}{\Sigma_{In}(t)} = 2\frac{1 - \exp\left(-2\frac{tc}{l_s}\right)}{1 + g_1^t(n)^2 - 2\exp\left(-2\frac{tc}{l_s}\right)}. \quad \text{(A10)}$$

Conti, S. G., and Demer, D. A. (**2003**). "Wide-bandwidth acoustical characterization of anchovy and sardine from reverberation measurements in an echoic tank," ICES J. Mar. Sci. **60**(3), 617–624.

Conti, S. G., Demer, D. A., and Brierley, A. S. (**2005**). "Broadbandwidth sound scattering and absorption from krill (*Meganyctiphanes norvegica*), Mysids (*Praunus flexuousus* and *Neomysis integer*) and shrimp (*Crangon crangon*)," ICES J. Mar. Sci. **62**(5), 656–965 .

Conti, S. G., Roux, P., Demer, D. A., and de Rosny, J. (**2004**). "Measurement of the scattering and absorption cross-sections of the human body," Appl. Phys. Lett. **84**, 819–821.

Cook, R., Waterhouse, R., Berendt, R., Edelman, S., and Thompson, M. (**1955**). "Measurements of correlation coefficients in reverberant sound fields," J. Acoust. Soc. Am. **27**, 1072–1077.

Cowan, M. L., Page, J. H., and Weitz, D. A. (**2000**). "Velocity fluctuations in fluidized suspensions probed by ultrasonic correlation spectroscopy," Phys. Rev. Lett. **85**, 453–456.

Cowan, M. L., Jones, I. P., Page, J. H., and Weitz, D. A. (**2002**). "Diffusing acoustic wave spectroscopy," Phys. Rev. E **65**, 066605.

de Rosny, J. (2000). "Milieux réverbérants et réversibilité," Doctoral Thesis, Université Paris VI, p. 156.

de Rosny, J., and Roux, P. (**2001**). "Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**, 2587–2597.

de Rosny, J., Roux, P., Fink, M., and Page, J. H. (**2003**). "Field fluctuation spectroscopy in a reverberant cavity with moving scatterers," Phys. Rev. Lett. **90**, 094302.

Demer, D. A., and Conti, S. G. (**2003**). "Validation of the stochastic distorted-wave Born approximation model with broad bandwidth total target strength measurements of Antarctic krill," ICES J. Mar. Sci. **60**(3), 625–635.

Demer, D. A., Conti, S. G., de Rosny, J., and Roux, P. (**2003**). "Absolute measurements of total target strength from reverberation in a cavity," J. Acoust. Soc. Am. **113**, 1387–1394.

Ishimaru, A. (**1978**). *Wave Propagation and Scattering in Random Media* Academic, New York, Vols. I and II.

Maret, G., and Wolf, P. E. (**1987**). "Multiple light-scattering from disordered media—the effect of Brownian motion of scatterers," Z. Phys. B: Condens. Matter **65**, 409–431.

Pine, D. J., Weitz, D. A., Chaikin, P. M., and Herbolzheimer, E. (**1988**). "Diffusing wave spectroscopy," Phys. Rev. Lett. **60**, 1134–1137.

Schroeder, M. (**1959**). "Measurement of sound diffusion in reverberation chambers," J. Acoust. Soc. Am. **31**, 1407–1414.

Sheng, P. (**1995**). *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, New York).

Viscous scattering of a pressure wave: Calculation of the fluid tractions on a biomimetic acoustic velocity sensor

Dorel Homentcovschi[a] and Ronald N. Miles
Department of Mechanical Engineering, SUNY Binghamton, New York, 13902-6000

(Received 28 June 2005; revised 1 November 2005; accepted 9 November 2005)

In the paper we give a method for calculating the tractions (local forces) of the fluid motion determined by an incoming plane pressure wave on an artificial hair cell transducer structure. The sensing element of the transducer is a standing high aspect ratio cilium in the shape of a narrow thin curved beam (tape-like), which can be easily fabricated in micro-/nanotechnology. The method is based on considering the system of partial differential equations describing the motion of the compressible viscous fluid in an acoustic linearized approximation, and representation of the velocity field as a viscous acoustic single-layer potential. The boundary conditions, stating the cancellation of the velocity components on the solid beam, yield a two-dimensional (2-D) system of three integral equations over the beam's surface for the traction components. In the case of a narrow cilium, the system of integral equations furnishes a system of two 1-D integral equations over the symmetry curve of the structure for obtaining the tangential and normal components of the traction. This system is solved numerically by a finite (boundary) element method. The numerical code written for solving the problem was applied to some particular structures. The last structure is similar to the trichobothrium of a spider *Cupiennius salei*. The results obtained show that the curvature of the hair is enhancing sensitivity to flows directed normal to the main shaft of the hair confirming the assertion of Barth *et al.* [Philos. Trans. R. Soc. London, Ser. B **340**, 445–461 (1993)]. [DOI: 10.1121/1.2146108]

PACS number(s): 43.20.Fn [TDM] Pages: 777–787

I. INTRODUCTION

Many insects can detect the low-velocity movement of the ambient air by means of hair sensilla that are deflected from the resting position by the air motion. The sensilla respond to the sound and wind as long as the frequency of the incoming signal is small. This is the case in crickets whose filiform cercal hairs vibrate in a sound field,[1] in caterpillars' that react to the airborne vibrations of an approaching predator by means of filiform hairs on the thorax,[2,3] in cockroaches and grasshoppers that can have thousands of filiform sensory hairs of various sizes used for detecting danger.[4] The spider's filiform hairs, also referred to as trichobothria, form spatial clusters and areas capable of detecting the magnitude, direction and frequency of airborne signals.[5,6]

Fish use lateral line sensors to monitor sounds under water.[7,8] The lateral line system consists of an array of distributed sensor nodes, each of them being a mechanoreceptor having as a basic element a vertical cilium attached to sensory cells. When the cilium of the hair cell is bent by the water flow, the displacement will induce output responses from the attached nerve cells.

Filiform hair systems attracted attention of several authors who derived and applied physical–mathematical models to describe the behavior of sensilla and their interaction with surrounding fluid motion. Substantial progress has been made in the understanding of the physics behind the working of individual sensory hairs. Shimozawa and Kanou[9] proposed a model of a hair as a slender linearly tapered cone. The viscous force acting on the hair shaft was obtained by using the Oseen's approximation for the drag force generated on a circular cylinder by a steady-state flow of a viscous fluid given by Imai.[10] A more realistic shape of hair receptors as elongated paraboloids was considered by Kumagai *et al.*[11]

Humphrey *et al.*[12] made an extensive critical examination of Shimozawa and Kanou's results[9] and proposed a mathematical model of the oscillatory motion of filiform hairs of all arthropods. The shape of the hair is assumed to be a straight, cylindrically shaped (rod-like) body of finite length and diameter. They used as the driving force on the hair shaft (driven by oscillating air motion) the same expression as the drag force generated on the rod of a solid pendulum swinging in stationary air, given by Stokes in 1851.[13] The Stokes's solution was obtained by solving the linearized system of Navier–Stokes equations (the Stokes's approximation). The numerical method developed by Humphrey *et al.*[12,14,15] solves a rigorously derived form of the equation for the conservation of angular momentum for a single hair; its estimation of the dynamic drag force is very informative.

The comparison of the measured properties of hair and air motion with the values predicted numerically by the theory developed by Barth *et al.*[16] showed very good agreement. The theoretical considerations are of a very general nature: they can be applied not only to different hair morphologies and hair mechanics but also to different media (air or water).[17–19]

[a]Permanent address: Polytechnica University, Applied Science Department & Institute of Mathematical Statistics and Applied Mathematics of Romanian Academy, Calea13 Septembrie #13, RO-76100, Bucharest, Romania. Electronic mail: homentco@binghamton.edu

The exhaustive analysis in Ref. 20 has shown that the use of the Stokes' approximation for computing the drag force gives better results than Oseen's model. The conclusion is that Oseen's approximation is not appropriate and is not used hereafter in approaching these problems.

In a recent physical and mathematical approach, Humphrey *et al.*[21] examined the relative importance of the various hair parameters in determining a hair's absolute sensitivity to medium flow as well as its frequency tuning. The effects of different parameters affecting the response of a hair-like medium flow detector are also given for hairs in water.[18,21,22]

As pointed out in Ref. 22, there are two major objectives of this work. The first is to uncover and understand the basic "design" principles underpinning the performance characteristics of filiform hairs by trying to understand how physics impacted the sensory ecology and adaptive evolution of the natural motion sensors. The second aim is to derive and implement realistic physical–mathematical models for these exquisitely sensitive natural sensors. A model that predicts the hair's response will expedite the design and fabrication of artificial sensors of similar function and characteristics. The value of the engineering approach to hair sensilla sensitive to medium flow is also underlined by the analysis by Shimozawa *et al.*[23] The conclusion in Barth *et al.*[24] is that the outstanding sensitivity seen in the neural response of cricket filiform hairs appear to represent the most sensitive biological sensors so far known.

A number of researchers have recognized the utility of the hair cell transducer structure and have applied MEMS techniques to produce microscale artificial hair cell sensors for sensing the perturbations of the air or water. [25–30] Microfabrication offers the benefits of high spatial resolution, fast time response, integrated signal processing and, finally, low costs.

We discuss the class of air- or water-sensing devices, based on the momentum transfer principle, using a vertical high aspect ratio cilium in the shape of a narrow thin curved beam (tape-like) which can be easily micromachined. This type of sensing element has not been found in nature, but the hope is that they can reproduce some of the hair's functions. The sensor's output is related to the direction and the intensity of the flow. By providing more than two sensors, with their cilia pointing in different directions, it is possible to identify the direction of the local air flow (or sound).[25] Practically, these artificial cell sensors are grouped in arrays of sensors with systematically varying frontal orientation and cilium shape.

In this paper we focus on the mechanical interaction of the air flow with individual hair-like sensors. When the motion of the ambient medium ranges in the domain of low velocities and the cross dimension of the body is smaller than the thickness of the viscous boundary layer of the supporting substrate, the force density acting on the hair (the traction) is dominated by the viscosity of the flow. The relation between the incoming sound or velocity field and the output of the sensor is obtained by first solving the equations of the motion of the viscous compressible fluid in the linear acoustic approximation with specified boundary conditions for the traction on the sensing element; the next step, which we are not considering here, is a structure deformation analysis under the known tractions. The model described here is tailored specially for determining the tractions (understood as local forces) on the tape-like sensors. The resultant forces (particularly the drag resulting by summing all the local forces) are strongly dependent on the geometry of the problem and are very different from that corresponding to rod-like sensilla found on insects. As a result, the numerical data obtained from this theory cannot be readily compared with those obtained in the above-cited papers. The validation of the theory can be done by comparing the theoretical results with experimental values obtained for artificial tape-like motion sensors. On the other hand, some qualitative results can be transferred between rod-like and tape-like sensors. Thus, the last example considered in this paper can be compared to results obtained for the trichobotrium of a spider *(Cupiennius salei)*. Our calculations support the assertion of Barth *et al.*[16] that the role of curvature on the sensing hairs is to enhance the sensitivity to flows directed normal to the main shaft of the hair.

Despite their simplicity, these "rudimentary" velocity sensors have an advantage over their very "sophisticated" inspiring natural sensors. While every natural hair shaft is a single sensing element (characterized by total drag force), for tape-like artificial sensors it is possible, depending on the detection technique, to obtain more data as local forces in different assigned points. This way the function of a cluster of natural hair sensors could be substituted by just a few artificial hair-sensing elements.

The solution of the linearized equations of viscous acoustics is developed as a single-layer viscous acoustical potential, which leads to a two-dimensional, regular Fredholm integral equation of the first kind for determining the tractions on the sensors' surface. Accounting for the fact that the beam is narrow, an asymptotic analysis of the integral equation yields a unidimensional integral equation. This technique is similar to that used to obtain the lifting line equation in classical aerodynamics.[32] This unidimensional integral equation is solved by a boundary (finite) element technique for the tractions in the direction of the normal (at the sensors' surface) and tangent to the sensors' middle curve. Based on these theoretical considerations a numerical code has been written and some results are provided in Sec. IV D.

II. EQUATIONS OF MOTION OF A VISCOUS COMPRESSIBLE FLUID IN LINEAR ACOUSTICS APPROXIMATION FOR HARMONIC OSCILLATIONS IN TIME

A. The equations of the motion of a viscous fluid in the linear acoustic approximation

If the coordinate system is chosen so that the unperturbed fluid is at rest, the first-order equations describing the isentropic flow of the gas can be written as[31,33,34]

$$\frac{1}{c_0^2}\frac{\partial}{\partial t}\frac{p'}{\rho_0} + \nabla \cdot \mathbf{v}' = 0, \tag{1}$$

$$\frac{\partial \mathbf{v}'}{\partial t} + \frac{1}{\rho_0}\nabla \cdot \sigma' = \mathbf{0}, \tag{2}$$

where p' and v' denote the pressure and velocity perturbations, σ' is the stress tensor that in the case of Newtonian fluids has the expression

$$\sigma'_{ij} \equiv \sigma_{ij}[p', \mathbf{v}'] = \left[p' - \left(\mu_B - \frac{2}{3}\mu \right) \nabla \cdot \mathbf{v}' \right] \delta_{ij} - \mu \left(\frac{\partial v'_i}{\partial x_j} + \frac{\partial v'_j}{\partial x_i} \right) \tag{3}$$

ρ_0, c_0 are the density and velocity of sound in a nonperturbed fluid, and by μ and μ_B we denote the shear and bulk viscosities.

The above equations are associated with the nonslip boundary condition

$$\mathbf{v}' = 0, \quad \text{on } \mathcal{S}, \tag{4}$$

where the solid surfaces $\mathcal{S}$ limit the flow domain $\mathcal{D}$. Thus, the principal element to be determined by solving a viscous acoustical problem is the velocity field (a vectorial unknown field).

B. The basic equation in the case of harmonic oscillations in time

We consider the case where all the physical variables are harmonic in time with the same angular velocity $\omega = 2\pi f$. The case of the general time dependence can be obtained, after analyzing each frequency separately, by Fourier superposition. In the case of simple harmonic oscillations in time we shall write

$$\{p'(\mathbf{x},t), \mathbf{v}'(\mathbf{x},t), \sigma'(\mathbf{x},t)\} = \{p(\mathbf{x}), \mathbf{v}(\mathbf{x}), \sigma(\mathbf{x})\} \exp(-i\omega t),$$

In this case, the continuity equation (1) becomes

$$\nabla \cdot \mathbf{v} = \frac{i\omega}{c_0^2}\frac{p}{\rho_0}. \tag{5}$$

Also, the momentum conservation equation can be written as

$$-i\omega \mathbf{v} + \frac{1}{\rho_0}\nabla \cdot \sigma = \mathbf{0}, \tag{6}$$

which in the case of Newtonian fluids becomes

$$\Delta \mathbf{v} + \frac{i\omega}{\nu}\mathbf{v} = \beta \nabla \frac{p}{\varrho_0}. \tag{7}$$

Here we have denoted

$$\beta = \frac{\rho_0 - (\mu/3 + \mu_B) i\omega / c_0^2}{\mu}.$$

The relationships (5) and (7) give the equation for the pressure,

$$[\Delta + k^2] p = 0, \tag{8}$$

which in the case $\mu = \mu_B = 0$ coincides with the basic equation determining the motion of the inviscid compressible fluid in the linear acoustic approximation. Finally, by applying the operator of Eq. (8) to Eq. (7), there results the basic equation describing the motion of the viscous compressible fluid in linear acoustic approximation,

$$[\Delta + k^2][\Delta + k^{*2}]\mathbf{v} = \mathbf{0}. \tag{9}$$

Here we have used the notations

$$k = \frac{\omega}{\sqrt{c_0^2 - i\omega(4\mu/3 + \mu_B)/\rho_0}}, \quad k^* = \sqrt{\frac{i\omega\rho_0}{\mu}},$$
$$\mathrm{Im}(k, k^*) \geqslant 0.$$

It is to be noticed that Eq. (8) is a Helmholtz-type equation and the operator in Eq. (9) is a product of two Helmholtz-type operators. Consequently, the velocity can be written as a sum of two terms: the first describes a propagation mode (also called the acoustical mode) and the second is a diffusion mode driven by viscosity.

C. Plane wave solution in the whole space

Consider an incoming pressure plane wave,

$$p^{\mathrm{in}}(\mathbf{x}) = \rho_0 c_0^2 P_0 \exp\{ik\mathbf{n} \cdot \mathbf{x}\}, \tag{10}$$

where $\mathbf{n}$ is the unit vector of the propagating direction and P_0 is a dimensionless constant used for scaling the amplitude of the incoming wave. It can be verified directly that (10) satisfies the basic pressure equation. Since the pressure field has an assigned form, Eq. (7) will determine the associated velocity field as

$$\mathbf{v}^{\mathrm{in}}(\mathbf{x}) = ik\delta c_0^2 p_0 \mathbf{n} \exp\{ik\mathbf{n} \cdot \mathbf{x}\}, \tag{11}$$

where

$$\delta = \frac{\rho_0 - (4\mu/3 + \mu_B) i\omega / c_0^2}{i\omega\rho_0 - \mu k^2}.$$

We notice that in this case that the velocity field contains only a propagating mode.

D. Plane wave solution in the half-space

Let us consider now the domain $\mathcal{D}$ as being the upper half-space $z > 0$. The $z = 0$ plane is assumed a solid boundary, hence the solution has to satisfy the no-slip boundary condition

$$\mathbf{v}(x, y, 0) = 0. \tag{12}$$

In the case the incident wave has the form

$$p^{\mathrm{in}}(\mathbf{x}) = \rho_0 c_0^2 P_0 \exp\{ik\mathbf{n} \cdot \mathbf{x}\}, \quad n_z \neq 0; \tag{13}$$

we consider the pressure field of the form

$$p(\mathbf{x}) = \rho_0 c_0^2 P_0 [\exp\{ik\mathbf{n} \cdot \mathbf{x}\} + A \exp\{ik\mathbf{n}' \cdot \mathbf{x}\}], \tag{14}$$

where $\mathbf{n} = (n_x, n_y, n_z)$, $\mathbf{n}' = (n_x, n_y, -n_z)$, and A is a constant. By using formula (11) we can write

$$\mathbf{v}(\mathbf{x}) = ik\delta c_0^2 P_0 [\mathbf{n} \exp\{ik\mathbf{n} \cdot \mathbf{x}\} + A\mathbf{n}' \exp\{ik\mathbf{n}' \cdot \mathbf{x}\}] + \mathbf{u}^0 c_0^2 P_0 \exp[ik(n_x x + n_y y) - qz], \tag{15}$$

where

$$q = \sqrt{k^2(n_x^2 + n_y^2) - k^{*2}}, \quad \mathrm{Re}(q) > 0.$$

The constants A and $\mathbf{u}^0$ are determined such that the solution satisfies the nonslip boundary condition:

$$A = \frac{qn_z - ik(n_x^2 + n_y^2)}{qn_z + ik(n_x^2 + n_y^2)},$$

$$u_x^0 = \frac{-2ikq\,\delta n_x n_z}{qn_z + ik(n_x^2 + n_y^2)}, \quad u_y^0 = \frac{-2ikq\,\delta n_y n_z}{qn_z + ik(n_x^2 + n_y^2)},$$

$$u_z^0 = \frac{2k^2\delta(n_x^2 + n_y^2)n_z}{qn_z + ik(n_x^2 + n_y^2)}. \tag{16}$$

It should be noted that the complementary velocity $\mathbf{u} = \mathbf{u}^0 c_0^2 P_0 \exp[ik(n_x x + n_y y) - qz]$ has a boundary layer structure.

In the solution (14), (15) one can recognize the incident wave (the first term in brackets), the reflected wave (the second term), and the contribution of the viscous boundary layer (the last term). Consequently, the solution in this case contains a propagating mode and also a diffusive (viscous) mode. Also, relation (16) yields the refection coefficient for the acoustic pressure. The solution of the problem of reflection of a transverse wave from a flat boundary in the case of an incompressible fluid can be found in Ref. 35. The most important property of the reflected wave is that its amplitude decreases exponentially as the distance from the solid surface increases.

III. THE FUNDAMENTAL FORMULA AND THE BOUNDARY INTEGRAL EQUATION FOR THE MOTION OF A VISCOUS COMPRESSIBLE FLUID IN THE LINEAR ACOUSTIC APPROXIMATION

A. The fundamental formula

We assume that the body occupying the domain $\mathcal{D}^+$ limited by the surface $\mathcal{S}$ is immersed into an external flow field characterized by the pressure $p^0(\mathbf{x})$ and velocity $\mathbf{v}^0(\mathbf{x})$ that are solutions of the equations of the linearized viscous acoustics. These functions can be the plane wave solution in the whole space or that corresponding to the half-space $z>0$. The solution of the problem $[p(\mathbf{x}), \mathbf{v}(\mathbf{x})]$ is defined in the external domain $\mathcal{D}^-$.

It is possible to use a direct approach to solve the boundary-value problem by means of the finite difference or finite element methods. Due to the infinity of the domain these equations will be written for all the mesh points inside domain $\mathcal{D}^-$. We prefer a boundary integral formulation that involves more mathematics but results in a much simpler system of equations written only on the mesh points on the sensor's surface. The starting elements are the following integral relationships that are proved in the Appendix:

$$\gamma p(\mathbf{x}) = p^0(\mathbf{x}) - \boldsymbol{\nabla}\cdot\int\int_{\mathcal{S}} \frac{\mathbf{t}(\mathbf{x}')\exp(ik|\mathbf{x}-\mathbf{x}'|)}{4\pi(1 - i\omega\nu'/c_0^2)|\mathbf{x}-\mathbf{x}'|}da',$$

$$\gamma\mathbf{v}(\mathbf{x}) = \mathbf{v}^0(\mathbf{x}) + \int\int_{\mathcal{S}} \frac{\mathbf{t}(\mathbf{x}')\exp(ik^*|\mathbf{x}-\mathbf{x}'|)}{4\pi\rho_0\nu|\mathbf{x}-\mathbf{x}'|} \times da' + \boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\int\int_{\mathcal{S}} \times \frac{\mathbf{t}(\mathbf{x}')[\exp(ik^*|\mathbf{x}-\mathbf{x}'|) - \exp(ik|\mathbf{x}-\mathbf{x}'|)]}{4\pi i\omega\rho_0|\mathbf{x}-\mathbf{x}'|}da'. \tag{17}$$

By da' we denoted the surface area element with respect to variable $\mathbf{x}'$ on the surface $\mathcal{S}$. Once the traction $\mathbf{t}(\mathbf{x}')$ on the boundary surface $\mathcal{S}$ is known, these relationships enable us to determine the velocity and pressure in any point in the domain $\mathcal{D}^-$. Since in the case of viscous fluids the boundary conditions on hard bodies are expressed in terms of velocities, the relationship (17) will be used more often. This is why we call this equation *the fundamental integral formula*.

Remark 1: In obtaining the relationship (17) *we have used the condition of no slip of the fluid along the surface* $\mathcal{S}$ (4). *This is why we have in formula* (17) *only a single-layer viscous acoustic potential*.

The fundamental formula can also be written as

$$\gamma\rho_0 i\omega\mathbf{v}(\mathbf{x}) = \rho_0 i\omega\mathbf{v}^0(\mathbf{x}) + \int\int_{\mathcal{S}} \frac{\mathbf{t}(\mathbf{x}')}{4\pi|\mathbf{x}-\mathbf{x}'|} \times A(|\mathbf{x}-\mathbf{x}'|,k,k^*)da' + \int\int_{\mathcal{S}} \frac{(\mathbf{x}-\mathbf{x}')[\mathbf{t}(\mathbf{x}')\cdot(\mathbf{x}-\mathbf{x}')]}{4\pi|\mathbf{x}-\mathbf{x}'|^3} \times C(|\mathbf{x}-\mathbf{x}'|,k,k^*)da',$$

where

$$A(|\mathbf{x}|,k,k^*) = k^{*2}\exp(ik^*|\mathbf{x}|) - \frac{\exp(ik^*|\mathbf{x}|)}{|\mathbf{x}|^2}(1 - ik^*|\mathbf{x}|) + \frac{\exp(ik|\mathbf{x}|)}{|\mathbf{x}|^2}(1 - ik|\mathbf{x}|), \tag{18}$$

$$C(|\mathbf{x}|,k,k^*) = \frac{3\exp(ik^*|\mathbf{x}|)}{|\mathbf{x}|^2}\left(1 - ik^*|\mathbf{x}| - \frac{k^{*2}|\mathbf{x}|^2}{3}\right) - \frac{3\exp(ik|\mathbf{x}|)}{|\mathbf{x}|^2}\left(1 - ik|\mathbf{x}| - \frac{k^2|\mathbf{x}|^2}{3}\right). \tag{19}$$

The behavior of the functions A and C for small values of $|\mathbf{x}|$ is given by formulas

$$A(|\mathbf{x}|,k,k^*)=\frac{1}{2}(k^2+k^{*2})+\frac{i|\mathbf{x}|}{3}(k^3+2k^{*3})+O[(k^4+k^{*4})|\mathbf{x}|^2],$$

$$C(|\mathbf{x}|,k,k^*)=\frac{1}{2}(k^{*2}-k^2)+O[(k^4+k^{*4})\times|\mathbf{x}|^2].$$

B. The boundary integral equation of the problem

By taking in the *fundamental integral formula* $\mathbf{x}=\mathbf{x}_0 \in \mathcal{S}$ there results

$$\int\int_{\mathcal{S}}\frac{(\mathbf{x}_0-\mathbf{x}')[\mathbf{t}(\mathbf{x}')\cdot(\mathbf{x}_0-\mathbf{x}')]}{4\pi|\mathbf{x}_0-\mathbf{x}'|^3}C(|\mathbf{x}_0-\mathbf{x}'|,k,k^*)da'+\int\int_{\mathcal{S}}\frac{\mathbf{t}(\mathbf{x}')}{4\pi|\mathbf{x}_0-\mathbf{x}'|}A(|\mathbf{x}_0-\mathbf{x}'|,k,k^*)da'+i\omega\rho_0\mathbf{v}^0(\mathbf{x}_0)=0. \tag{20}$$

This is a Fredholm integral equation of the first kind for determining the boundary traction $\mathbf{t}(\mathbf{x})$. Analytical solutions of this equation can be obtained only in very particular cases. Therefore, it has to be solved numerically by boundary-element-type methods.

It is clear that the kernel of (20) has a singularity for $\mathbf{x}'=\mathbf{x}_0$. The singular part of the integral operator can be written as

$$\mathbb{I}[\mathbf{t}]=\frac{1}{2}(k^2+k^{*2})\int\int_{\mathcal{S}}\frac{\mathbf{t}(\mathbf{x}')}{|\mathbf{x}_0-\mathbf{x}'|}da'+\frac{1}{2}(k^{*2}-k^2)\int\int_{\mathcal{S}}\frac{(\mathbf{x}_0-\mathbf{x}')[\mathbf{t}'\cdot(\mathbf{x}_0-\mathbf{x}')]}{|\mathbf{x}_0-\mathbf{x}'|^3}da'.$$

This form shows that when $\mathcal{S}$ is a Lyapunov surface, the kernels are weakly singular. For $\omega=0$ the equation (20) reduces to the Oseen operator for the viscous incompressible flow in Stokes' approximation. Consequently, the methods used to approximate the solution in the case of viscous incompressible fluids[37] can be successfully applied for integrating Eq. (20).

IV. CALCULATION OF THE TRACTION ON A LOW-FREQUENCY SENSOR

A. The integral equation of the curvilinear sensor approximation

We consider now the case of a low-frequency sensor having the shape of a curved narrow thin beam, symmetrical with respect to the plane $x=0$, of length L and width $2d_0g(s)$ [$g(s)$ being a given function, $|g(s)|\leqslant 1$] such that $\varepsilon=d_0/L\ll 1$. The range of frequencies considered here is $0.1<f<500$ Hz. At these frequencies the wavelength is large compared with the length L. The equation of the surface $\mathcal{S}$ is assumed of the form

$$y=y(s),\quad 0\leqslant s\leqslant L,$$
$$z=z(s),\quad 0\leqslant s\leqslant L,$$
$$x=x,\quad -d_0g(s),\leqslant x\leqslant d_0g(s).$$

This shape of the sensing element was chosen since it can be obtained in micro-/nanofabrication technology. Denote by $\mathcal{C}$ the central curve of the surface (Fig. 1), supposed to be contained in the plane $x=0$, and s is the curvilinear abscissa along $\mathcal{C}$. We shall write the integral equation (20) for the points $\mathbf{x}_0$ along the curve $\mathcal{C}$.

Also, we introduce dimensionless independent variables taking d_0 as the reference length along the x axis and L for the other directions. We also denote by capital letters the vector components orthogonal to the $\hat{x}$ direction. (By $\hat{\mathbf{v}}$ we denote the unit vector corresponding to the direction of vector $\mathbf{v}$.) Thus, we have

$$\mathbf{x}\to L(\varepsilon x\hat{\mathbf{x}}+\mathbf{R}),\quad \mathbf{R}=(0,y/L,z/L),$$

$$|\mathbf{x}_0-\mathbf{x}'|\to L\{\varepsilon^2x'^2+|\mathbf{R}_0-\mathbf{R}'|^2\}^{1/2},$$

$$\mathbf{t}=t_x\hat{\mathbf{x}}+\mathbf{T},\quad \mathbf{v}^0=v_x^0\hat{\mathbf{x}}+\mathbf{V}^0,$$

$$da'=\varepsilon L^2\,dx'\,ds'.$$

Since the surface is narrow and the kernel of the integral equation is an even function with respect to x', the system of integral equations (20) is separable into an equation corresponding to the x direction,

$$v_x^0(\mathbf{R}_0)+\int\int_{\mathcal{S}}\frac{t_x(\mathbf{R}')}{4\pi\rho_0|\mathbf{x}_0-\mathbf{x}'|}A(|\mathbf{x}_0-\mathbf{x}'|,\omega,\nu)da'+O(L^2\varepsilon^3)=0, \tag{21}$$

and a vectorial equation (a system of two integral equations) for the other directions,

$$\mathbf{V}^0(\mathbf{R}_0)+\int\int_{\mathcal{S}}\frac{\mathbf{T}(\mathbf{R}')}{4\pi\rho_0|\mathbf{x}_0-\mathbf{x}'|}A(|\mathbf{x}_0-\mathbf{x}'|,\omega,\nu)da'+L^2\int\int_{\mathcal{S}}\frac{(\mathbf{R}_0-\mathbf{R}')[\mathbf{T}(\mathbf{R}')\cdot(\mathbf{R}_0-\mathbf{R}')]}{4\pi\rho_0|\mathbf{x}_0-\mathbf{x}'|^3}\times C(|\mathbf{x}_0-\mathbf{x}'|,\omega,\nu)da'+O(L^2\varepsilon^3)=\mathbf{0}. \tag{22}$$

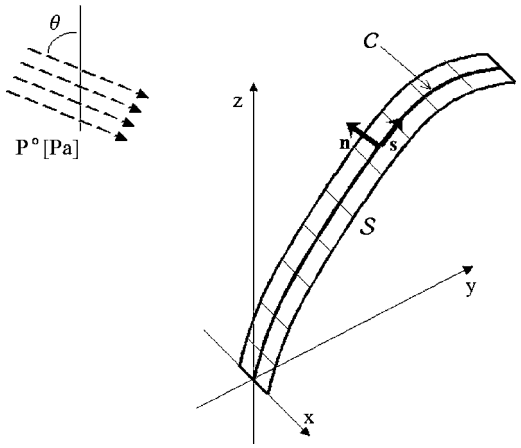


FIG. 1. The geometry of a sensing element.

We shall solve the system (22), which gives the main forces on the sensor surface. Let $F(\mathbf{x}')$ be a smooth function. We can write

$$F(\mathbf{x}') = F(\mathbf{R}') + x'\,\partial_1 F(\mathbf{R}') + O(\varepsilon^2),$$

and hence,

$$\int_{-d_0 g(s)}^{d_0 g(s)} \frac{F(\mathbf{x}')}{|\mathbf{x}_0 - \mathbf{x}'|} dx' = 2\int_0^{d_0 g(s)} \frac{F(\mathbf{R}')}{|\mathbf{x}_0 - \mathbf{x}'|} dx' + O(\varepsilon^3) = 2F(\mathbf{R}')\log\frac{\varepsilon g(s) + \sqrt{\varepsilon^2 g^2(s) + |\mathbf{R}_0 - \mathbf{R}'|^2}}{|\mathbf{R}_0 - \mathbf{R}'|} + O(\varepsilon^3),$$

$$L^2 \int_{-d_0 g(s)}^{d_0 g(s)} \frac{F(\mathbf{x}')}{|\mathbf{x}_0 - \mathbf{x}'|} dx' = \frac{2\varepsilon g(s) F(\mathbf{R}')}{|\mathbf{R}_0 - \mathbf{R}'|^2} \frac{1}{\sqrt{\varepsilon^2 g^2(s) + |\mathbf{R}_0 - \mathbf{R}'|^2}} + O(\varepsilon^3).$$

Taking into consideration these formulas, the integration with respect to x' in Eqs. (21) and (22) can be performed directly, and there results

$$2\pi\rho_0 \mathbf{V}^0(\mathbf{R}_0) + \int_C \mathbf{T}(\mathbf{R}') A(|\mathbf{R}_0 - \mathbf{R}'|) \times \log\frac{\varepsilon g(s)' + \sqrt{\varepsilon^2 g^2(s') + |\mathbf{R}_0 - \mathbf{R}'|^2}}{|\mathbf{R}_0 - \mathbf{R}'|} ds' + \varepsilon \int_C \frac{\mathbf{R}' - \mathbf{R}_0}{|\mathbf{R}' - \mathbf{R}_0|} \left(\mathbf{T}(\mathbf{R}') \cdot \frac{\mathbf{R}' - \mathbf{R}_0}{|\mathbf{R}' - \mathbf{R}_0|} \right) \times \frac{C(|\mathbf{R}_0 - \mathbf{R}'|)}{\sqrt{\varepsilon^2 g^2(s') + |\mathbf{R}_0 - \mathbf{R}'|^2}} g(s)' ds' = \mathbf{0}, \tag{23}$$

which is the integral equation for determining the traction on the low-frequency sensor.

Remark 2: In the above formulas it was assumed that the tractions are smooth functions across the beam. Equation (23) *still proves true in the case the tractions* $T(\mathbf{x}')$ *have integrable singularities* (*like square root singularities*) *at the beam's sharp edges but is a smooth function at all the other points.*

B. A boundary-element approach to the integral equation of the curvilinear sensor

For a numerical solution of Eq. (23) we consider the nodal points $\mathbf{R}_j^0$, $j=1,\ldots,N+1$ on the curve $\mathcal{C}$ and approximate the arc $\mathcal{C}_j$ between the points $\mathbf{R}_j^0$ and $\mathbf{R}_{j+1}^0$ of the curve by the line segment $\mathcal{E}_j \equiv \overline{\mathbf{R}_j^0 \mathbf{R}_{j+1}^0}$, which is the basic boundary element we will use. The length of this line segment is denoted by h_j. We shall determine the solution of the equation in the points $\mathbf{R}_j = (\mathbf{R}_j^0 + \mathbf{R}_{j+1}^0)/2$ lying at the middle of the boundary elements. Also, for the vectorial quantities (like velocities and tractions) we consider local Cartesian systems of coordinates given by the unit vector $\hat{\mathbf{s}}_j$ of the direction $\overrightarrow{\mathbf{R}_j^0 \mathbf{R}_{j+1}^0}$ and the normal unit vector $\hat{\mathbf{n}}_j$. We consider the traction $\mathbf{T}(\mathbf{R}')$ as constant along the boundary element $\mathcal{E}_j$ and equal to the value

$$\mathbf{T}_j = \mathbf{T}(\mathbf{R}_j) = T_j^s \hat{\mathbf{s}}_j + T_j^n \hat{\mathbf{n}}_j,$$

and will write the equation (23) in the points $\mathbf{R}_k$ projected on a local system of coordinates in the form

$$\sum_{j=1}^N M_{kj}^{nn} T_j^n + \sum_{j=1}^N M_{kj}^{ns} T_j^s = -2\pi i\omega\rho_0 V_k^{0n}/\varepsilon, \tag{24}$$

$$\sum_{j=1}^N M_{kj}^{sn} T_j^n + \sum_{j=1}^N M_{kj}^{ss} T_j^s = -2\pi i\omega\rho_0 V_k^{0s}/\varepsilon, \tag{25}$$

where

$$M_{kj}^{nn} = \hat{\mathbf{n}}_k \cdot \hat{\mathbf{n}}_j I_{kj} + J_{kj}[\hat{\mathbf{n}}_k, \hat{\mathbf{n}}_j], \quad M_{kj}^{ns} = [\hat{\mathbf{n}}_k \cdot \hat{\mathbf{s}}_j I_{kj} + J_{kj}[\hat{\mathbf{n}}_k, \hat{\mathbf{s}}_j]], \quad j,k = 1,\ldots,N,$$

$$M_{kj}^{sn} = \hat{\mathbf{s}}_k \cdot \hat{\mathbf{n}}_j I_{kj} + J_{kj}[\hat{\mathbf{s}}_k, \hat{\mathbf{n}}_j], \quad M_{kj}^{ss} = \hat{\mathbf{s}}_k \cdot \hat{\mathbf{s}}_j I_{kj} + J_{kj}[\hat{\mathbf{s}}_k, \hat{\mathbf{s}}_j], \quad j,k = 1,\ldots,N.$$

Here we have denoted

$$I_{kj} = \frac{1}{\varepsilon} A(|\mathbf{R}_j - \mathbf{R}_k|) \int \log\frac{\varepsilon g_j + \sqrt{\varepsilon^2 g_j^2 + |\mathbf{R}' - \mathbf{R}_k|^2}}{|\mathbf{R}' - \mathbf{R}_k|} ds', \tag{26}$$

$$\mathbf{J}_{kj}[\mathbf{a}_j] = C(|\mathbf{R}_j - \mathbf{R}_k|) g_j \int \frac{\mathbf{R}' - \mathbf{R}_k}{|\mathbf{R}' - \mathbf{R}_k|} \left(\frac{\mathbf{R}' - \mathbf{R}_k}{|\mathbf{R}' - \mathbf{R}_k|} \cdot \mathbf{a}_j \right) \times \frac{ds'}{\sqrt{\varepsilon^2 g_j^2 + |\mathbf{R}' - \mathbf{R}_k|^2}} \tag{27}$$

$$J_{kj}[\hat{\mathbf{s}}_k, \hat{\mathbf{n}}_j] = \hat{\mathbf{s}}_k \cdot \mathbf{J}_{kj}[\hat{\mathbf{n}}_j]; \quad J_{kj}[\hat{\mathbf{n}}_k, \hat{\mathbf{n}}_j] = \hat{\mathbf{n}}_k \cdot \mathbf{J}_{kj}[\hat{\mathbf{n}}_j],$$

$$J_{kj}[\hat{\mathbf{s}}_k, \hat{\mathbf{s}}_j] = \hat{\mathbf{s}}_k \cdot \mathbf{J}_{kj}[\hat{\mathbf{s}}_j]; \quad J_{kj}[\hat{\mathbf{n}}_k, \hat{\mathbf{s}}_j] = \hat{\mathbf{n}}_k \cdot \mathbf{J}_{kj}[\hat{\mathbf{s}}_j].$$

1. Calculation of the integrals I_{kj}, J_{kj} for $k \neq j$

To compute the integrals (26) and (27) for $j \neq k$, we use an asymptotic expansion around $\varepsilon = 0$ and evaluate analytically the resulting integrals. Neglecting $O(\varepsilon^2)$ terms there results

$$I_{kj} = A(|\mathbf{R}_j - \mathbf{R}_k|) g_j I_1,$$

$$\mathbf{J}_{kj}[\mathbf{a}_j] = C(|\mathbf{R}_j - \mathbf{R}_k|) g_j [a_j^t \hat{\mathbf{s}}_j I_1 + R_{jk}^n (a_j^n \hat{\mathbf{s}}_j + a_j^t \hat{\mathbf{n}}_j) I_2 + (-a_j^t \hat{\mathbf{s}}_j + a_j^n \hat{\mathbf{n}}_j) I_3].$$

Here we have used the notations

$$R_{jk}^t = (\mathbf{R}_j - \mathbf{R}_k) \cdot \hat{\mathbf{s}}_j, \quad R_{jk}^n = (\mathbf{R}_j - \mathbf{R}_k) \cdot \hat{\mathbf{n}}_j,$$
$$\mathbf{a}_j = a_j^t \hat{\mathbf{s}}_j + a_j^n \hat{\mathbf{n}}_j.$$

The integrals I_1, I_2, I_3 have the expressions

$$I_1 = \log \frac{R^+ + R_{jk}^t + h_j/2}{R^- + R_{jk}^t - h_j/2},$$

$$I_2 = \frac{2R_{jk}^t h_j}{R^+ R^- (R^+ + R^-)},$$

$$I_3 = \frac{h_j}{R^+ R^- (R^+ + R^-)} \left((R_{jk}^n)^2 - (R_{jk}^t)^2 + \left(\frac{h_j}{2} \right)^2 + R^+ R^- \right),$$

$$R^+ = \left| \sqrt{(R_{jk}^t + h_j/2)^2 + (R_{jk}^n)^2} \right|,$$

$$R^- = \left| \sqrt{(R_{jk}^t - h_j/2)^2 + (R_{jk}^n)^2} \right|.$$

2. Calculation of the integrals I_{kk}, J_{kk}

By performing the change of variables,

$$\mathbf{R}_k' = \mathbf{R}_k^0 + \left(\frac{1}{2} + \frac{t}{h_k} \right) (\mathbf{R}_{k+1}^0 - \mathbf{R}_k^0),$$

the integrals I_{kk} and J_{kk} become

$$I_{kk} = \frac{A(0)}{\varepsilon} \int_{-h_k/2}^{h_k/2} \log \frac{\varepsilon g_k + \sqrt{[\varepsilon g_k]^2 + t^2}}{|t|} dt = 2A(0) g_k \int_0^{h_k/(2\varepsilon g_k)} \log \frac{1 + \sqrt{1 + v^2}}{v} dv. \quad (28)$$

We have

$$J_{kk}[\hat{\mathbf{n}}_k, \hat{\mathbf{n}}_k] = J_{kk}[\hat{\mathbf{s}}_k, \hat{\mathbf{n}}_k] = J_{kk}[\hat{\mathbf{n}}_k, \hat{\mathbf{s}}_k] = 0$$

and

$$J_{kk}[\hat{\mathbf{s}}_k, \hat{\mathbf{s}}_k] = C(0) g_k \int_{-h_k/2}^{h_k/2} \frac{dt}{\sqrt{[\varepsilon g_k]^2 + t^2}} = 2C(0) g_k \log \left[\frac{h_k}{2\varepsilon g_k} + \sqrt{1 + \left(\frac{h_k}{2\varepsilon g_k} \right)^2} \right]. \quad (29)$$

The integral in formula (28) will be evaluated numerically.

Remark 3: *We note that in formulas* (28)–(29) *enters the ratio* h_k/ε. *In order to keep these terms finite, the number of elements has to be chosen such that this ratio is of order* $O(1)$. *This is the price paid for combining a small parameter asymptotic expansion with a finite element solution of the resulting integral equation.*

C. A boundary-element approach for the curvilinear sensor approximation in a half-space

In the case where the domain is the half-space $z>0$ we will consider also the symmetrical curve $\mathcal{C}'$ of $\mathcal{C}$ with respect to the plane $z=0$. In the case we have on the whole curvilinear arc $\mathcal{C}' \cup \mathcal{C}$ a number of $2N$ linear boundary elements we can write the equations (24) and (25) taking $2N$ instead of N. Taking also in consideration that for symmetrical elements we have

$$T_j^n = T_{2N+1-j}^n, \quad T_j^s = -T_{2N+1-j}^s, \quad j = 1, \ldots, N.$$

The system of equations for determining the tractions on the surface become

$$\sum_{j=1}^N Q_{kj}^{nn} T_j^n + \sum_{j=1}^N Q_{kj}^{ns} T_j^s = -2\pi i \omega \rho_0 V_k^{0n}/\varepsilon, \quad k = 1, \ldots, N, \quad (30)$$

$$\sum_{j=1}^N Q_{kj}^{sn} T_j^n + \sum_{j=1}^N Q_{kj}^{ss} T_j^s = -2\pi i \omega \rho_0 V_k^{0s}/\varepsilon, \quad k = 1, \ldots, N, \quad (31)$$

where

$$Q_{kj}^{nn} = M_{kj}^{nn} + M_{k,2N+1-j}^{nn}, \quad Q_{kj}^{ns} = M_{kj}^{ns} - M_{k,2N+1-j}^{ns}, \quad j = 1, \ldots, N,$$

$$Q_{kj}^{sn} = M_{kj}^{sn} + M_{k,2N+1-j}^{sn}, \quad Q_{kj}^{ss} = M_{kj}^{ss} - M_{k,2N+1-j}^{ss}, \quad j = 1, \ldots, N.$$

The traction on the surface is determined by the values T_j^n, $T_j^s (j=1, \ldots, N)$ resulting by solving the system (30), (31).

Remark 4: *We note that by taking also the symmetrical curve* $\mathcal{C}'$ *of* $\mathcal{C}$ *with respect to the* $z=0$ *plane we assure the cancellation of the normal velocity component along the z plane. The other two components are different from zero, but we expect them to be small due to the small perturbations produced by the sensor in the external domain. A complete solution has to consider also a traction (mechanical resistance) distribution along the plane* $z=0$. *But, the frictional resistance within the hair base can be assumed zero since, according to Ref.* 20, *no practical method is available to measure such slight mechanical resistance.*

D. Application

The developed theory and formulas were implemented into a MATLAB program. Various parameters for the sound field and the description of the shape and dimensions of the sound sensors can be input into the program. The tractions at the specified points on the central curve will be the output. These output forces can be utilized in a finite element model to gauge the harmonic response of the sound sensor.

In all the applications, the constant P_0 has been chosen such that the incoming pressure wave has the amplitude 1 Pa; the frequency is assumed to be 100 Hz.

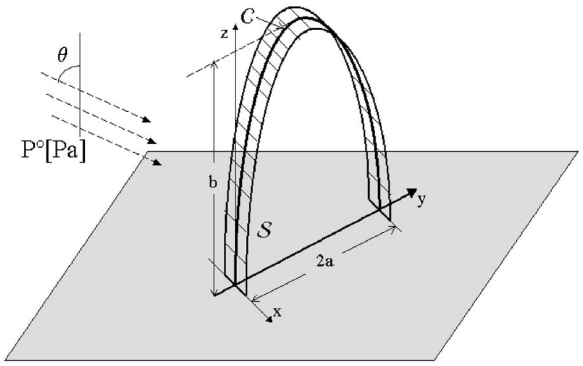


FIG. 2. A sensing element in the shape of a half-ellipse.

The program was tested on a half-elliptical shape sensor of 1 and 3.5 mm half-axis (Fig. 2) since it has two ends on the plane $z=0$ and a simple analytical expression. The width of the thin beam was chosen as 0.1 mm. The MATLAB program was run for three inclinations of the incident sound wave: $\theta=\pi/3,\pi/4,\pi/8$. The plots of the normal tractions are given in Fig. 3 and those of tangential tractions in Fig. 4. The continuous lines correspond to the real part of the tractions and the dotted line to the imaginary parts.

The next application contains the case of a vertical plane beam of variable width and a horizontal incoming plane wave ($\theta=\pi/2$) given in Fig. 5. Two cases were considered (1) ($a_1=0.01$ mm, $a_2=0.1$ mm) corresponding to a long trapezoid having its small basis in the plane $z=0$ and (2) ($a_1=0.1$ mm, $a_2=0.01$ mm) when the trapezoid has its big basis in the plane $z=0$. The moment of the forces resulting from normal tractions with respect to the $\mathcal{O}x$ axis are plotted in Fig. 6. The total moment is in the first case $MT_1=9.0\times10^{-8}-6.6\times10^{-9}i$(mm·N), and in the second case $MT_2=6.4\times10^{-8}-3.7\times10^{-9}i$(mm·N), which shows a strong influence of the shape of the thin beam on the moment. This is, in fact, the result of the action of the boundary layer on the plane $x\mathcal{O}y$. (The geometrical parameters of the beam in this case are: $L=1$ mm, *big basis* = 100 μm, *small basis* = 10 μm). Hence, for designing an artificial hair-cell sensor the first case would be preferable.

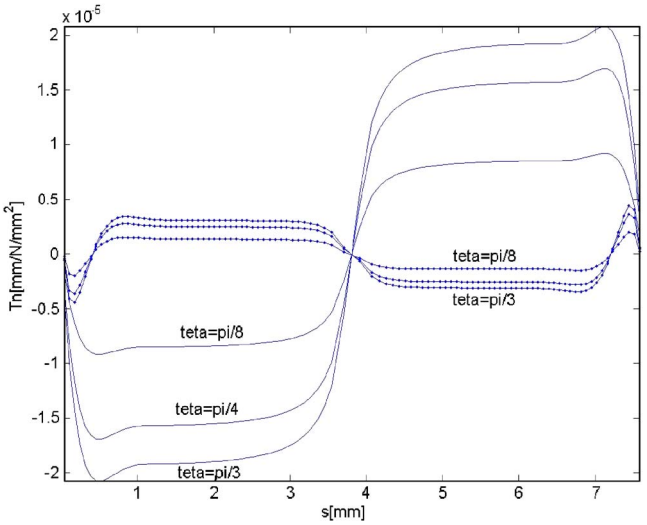


FIG. 3. The normal tractions for the sensing element in Fig. 2.

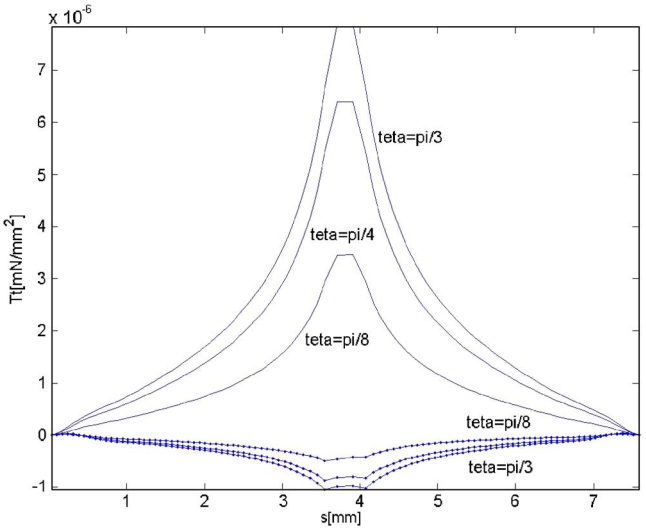


FIG. 4. The tangential tractions for the sensing element in Fig. 2.

Finally, the last application is that in Fig. 7, where we have a vertical rectangular narrow beam of length b that has attached at the upper end a horizontal rectangular beam of length a of the same width as the vertical beam. These beams are connected by a quarter of circle of radius r to ensure that the entire surface $\mathcal{S}$ is smooth. This structure is under the influence of a horizontal plane pressure wave of 1 Pa amplitude given by formula (15). For the case $a=8$ mm, $b=3$ mm, $r=0.1$ mm, and $width=0.14$ mm, the tractions on the structure are plotted in Fig. 8 (the real part) and in Fig. 9—the imaginary part. This structure is also advantageous for designing artificial hair-like sensors since all the force resulting from the tractions on the horizontal beam are acting on the upper end of the vertical beam, which will give a larger deflection of this component. This example is very similar to the trichobothrium of a spider *Cupiennius salei*. The curvature of the motion sensing hair is enhancing the sensitivity of the trichobothrium to flows normal to the main shaft of the hair. This fact supports the speculation by Barth *et al.* concerning the role of curvature in the sensing process.[16]

By reversing the direction of the incoming pressure wave the tractions in the last application remain unchanged. This can be considered as a result of the theorem (of Olmstead and Gautesen [38]) concerning the drag invariance for the

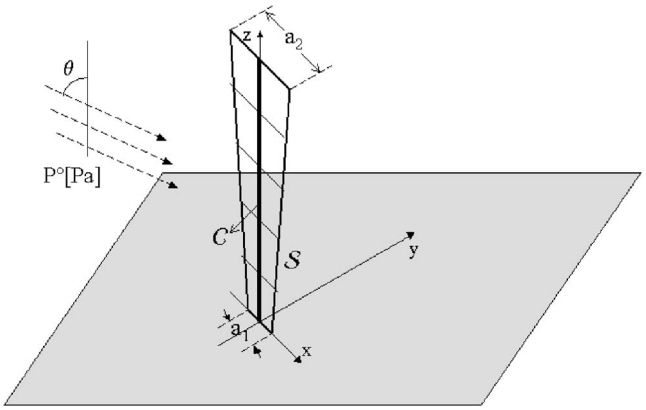


FIG. 5. The sensing element in the shape of an erected trapezoid.

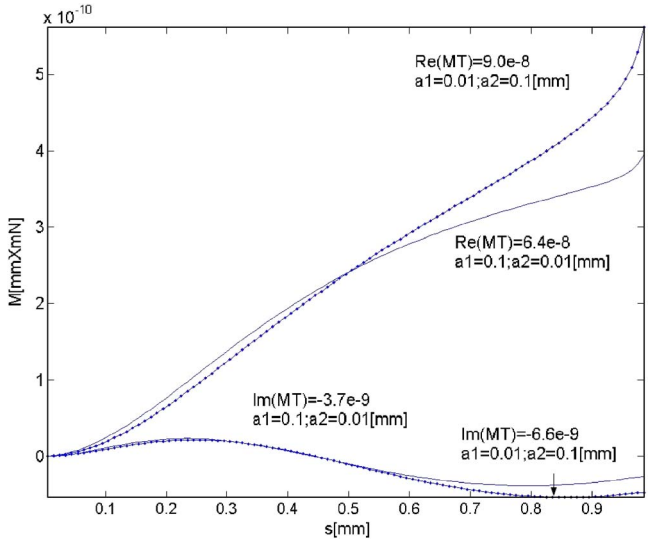


FIG. 6. The moment of normal tractions for the sensing element in Fig. 5 for two different geometries: the trapezoid having the small basis on the plane $z=0$ (dotted line) and the case the large basis is on the plane $z=0$ (simple continuous line).

reversal of the flow velocity for an arbitrary body. This paradox, proved in the case of incompressible flow, is still valid for the flow of a compressible viscous fluid in the linearized acoustic approximation.

In all the considered examples, the imaginary part of the tractions are much smaller than the real part, showing a strong influence of the viscosity.

ACKNOWLEDGMENTS

The careful, constructive suggestions of the reviewers is greatfully acknowledged. This work has been supported through a NIH Grant No. R01 DC05762-1A1 to RNM.

APPENDIX: THE FUNDAMENTAL INTEGRAL FORMULA

We shall assume that the body occupying the domain $\mathcal{D}^+$ limited by the surface $\mathcal{S}$ is immersed into an external flow field characterized by the pressure $p^0(\mathbf{x})$ and velocity $\mathbf{v}^0(\mathbf{x})$, which are satisfying the basic equations. The solution of the

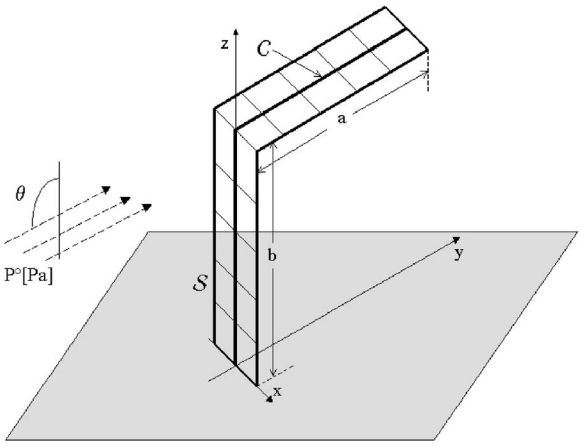


FIG. 7. The sensing element in the shape of a horizontal narrow beam attached at the upper end of a vertical beam.

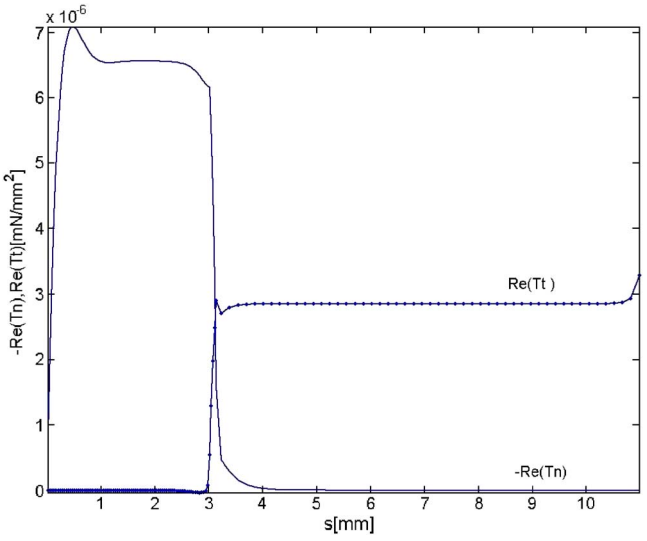


FIG. 8. The real part of normal tractions (continuous line) and the tangential tractions (dotted line) for the geometry in Fig. 7.

problem $[p(\mathbf{x}),\mathbf{v}(\mathbf{x})]$ is defined in the external domain $\mathcal{D}^-$. We extend these functions with the value 0 in domain $\mathcal{D}^+$. Then we have

$$\frac{-i\omega}{c_0^2}\frac{p^0}{\rho_0}+\nabla\cdot\mathbf{v}^0=0,\quad -i\omega\mathbf{v}^0+\frac{1}{\rho_0}\nabla\cdot\ \sigma^0=\mathbf{0},\quad in\ \mathcal{D}^+\cup\mathcal{S}\cup\mathcal{D}^-, \tag{A1}$$

$$\frac{-i\omega}{c_0^2}\frac{p}{\rho_0}+\nabla\cdot\mathbf{v}^0=0,\quad -i\omega\mathbf{v}+\frac{1}{\rho_0}\nabla\cdot\sigma=\mathbf{0},\quad in\ \mathcal{D}^+\cup\mathcal{D}^-. \tag{A2}$$

Also, we define

$$p^*=p-p^0,\quad \mathbf{v}^*=\mathbf{v}-\mathbf{v}^0,\quad in\ \mathcal{D}^+\cup\mathcal{D}^-, \tag{A3}$$

which, by means of relationships (A1), (A2) are satisfying the equations

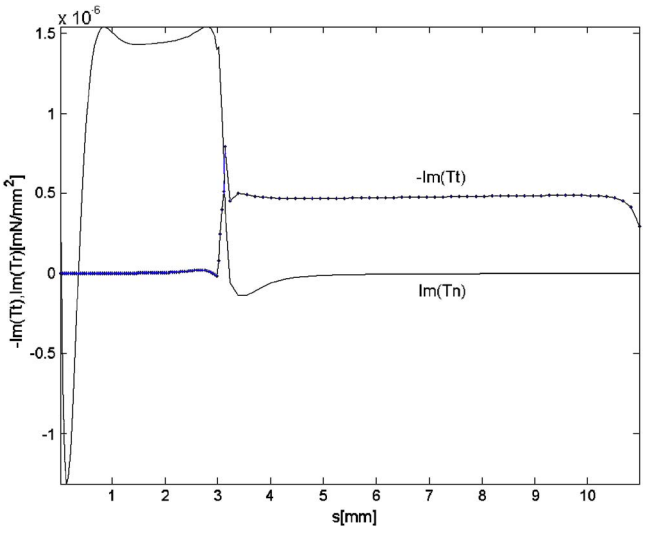


FIG. 9. The imaginary part of normal tractions (continuous line) and the tangential tractions (dotted line) for the geometry in Fig. 7.

$$\frac{-i\omega p^*}{c_0^2}+\frac{1}{\rho_0}\nabla\cdot\mathbf{v}^*=0,\quad -i\omega\mathbf{v}^*+\frac{1}{\rho_0}\nabla\cdot\sigma^*=\mathbf{0},\tag{A4}$$

in $\mathcal{D}^+\cup\mathcal{D}^-$.

Now, we consider the Fourier transform with respect to space variables,

$$\{\widetilde{p^*}(\mathbf{k}),\widetilde{\mathbf{v}^*}(\mathbf{k}),\widetilde{\sigma^*}(\mathbf{k})\}=\int\int\int\{p^*(\mathbf{x}),\mathbf{v}^*(\mathbf{x}),\sigma^*(\mathbf{x})\}\times\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x;$$

which results in

$$\widetilde{\nabla\cdot\sigma^*}=\int\int\int\nabla\cdot\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x=\int\int\int_{\mathcal{D}^+}\nabla\cdot\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x+\int\int\int_{\mathcal{D}^-}\nabla\cdot\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x.$$

But

$$\nabla\cdot\sigma^*\exp(-i\mathbf{k}\cdot\mathbf{x})=\nabla\cdot[\sigma^*\exp(-i\mathbf{k}\cdot\mathbf{x})]-[\nabla\exp(-i\mathbf{k}\cdot\mathbf{x})]\cdot\sigma^*=\nabla\cdot[\sigma^*\exp(-i\mathbf{k}\cdot\mathbf{x})]+i\mathbf{k}\cdot\sigma^*\exp(-i\mathbf{k}\cdot\mathbf{x}).$$

Then, we obtain

$$\int\int\int_{\mathcal{D}^+}\nabla\cdot\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x=i\mathbf{k}\cdot\int\int\int_{\mathcal{D}^+}\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x-\int\int_{\mathcal{S}}\mathbf{n}\cdot\sigma^0(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})da,\tag{A5}$$

$$\int\int\int_{\mathcal{D}^-}\nabla\cdot\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x=i\mathbf{k}\cdot\int\int\int_{\mathcal{D}^-}\sigma^*(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})dv_x-\int\int_{\mathcal{S}}\mathbf{n}\cdot[\sigma(\mathbf{x})-\sigma^0(\mathbf{x})]\exp(-i\mathbf{k}\cdot\mathbf{x})da,\tag{A6}$$

where $\mathbf{n}$ denotes the unit normal vector at the surface $\mathcal{S}$ pointing outward. The sum of relationships (A5) and (A6) gives

$$\widetilde{\nabla\cdot\sigma^*}=-i\mathbf{k}\cdot\widetilde{\sigma^*}-\int\int_{\mathcal{S}}\mathbf{t}(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})da,$$

where

$$\mathbf{t}(\mathbf{x})=\mathbf{n}\cdot\sigma(\mathbf{x})$$

is the traction (surface stress) at the surface. For the velocity we obtain similarly

$$\widetilde{\nabla\cdot\sigma^*}=-i\mathbf{k}\cdot\widetilde{\mathbf{v}^*},$$

where the condition $\mathbf{v}=\mathbf{0}$ on the solid surface $\mathcal{S}$ has been used.

The Fourier transform of the system (A4) gives

$$\frac{-i\omega\widetilde{p^*}}{c_0^2}+\frac{\widetilde{p^*}}{\rho_0}+i\mathbf{k}\cdot\widetilde{\mathbf{v}^*}=0,\tag{A7}$$

$$-i\omega\widetilde{\mathbf{v}^*}+i\mathbf{k}\cdot\frac{1}{\rho_0}\widetilde{\sigma^*}=\frac{1}{\rho_0}\int\int_{\mathcal{S}}\mathbf{t}(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})da.\tag{A8}$$

Thus, the algebraic system obtained in Fourier Transform space contains also on the right-hand side of Eq. (A8) the action of the surface $\mathcal{S}$ on the fluid motion. Since we have

$$\widetilde{\sigma_{ij}^*}=\left[\widetilde{p^*}-\left(\mu_B-\frac{2}{3}\mu\right)i\mathbf{k}\cdot\widetilde{\mathbf{v}^*}\right]\delta_{ij}-\mu(ik_j\widetilde{v_i^*}+ik_i\widetilde{v_j^*}),$$

the equation (A8) can be written in the form

$$(\nu|\mathbf{k}|^2-i\omega)\widetilde{\mathbf{v}^*}+i\mathbf{k}\cdot\left(\frac{\widetilde{p^*}}{\rho_0}-(\nu'-\nu)i\mathbf{k}\cdot\widetilde{\mathbf{v}^*}\right)=\frac{1}{\rho_0}\int\int_{\mathcal{S}}\mathbf{t}(\mathbf{x})\exp(-i\mathbf{k}\cdot\mathbf{x})da.\tag{A9}$$

The inner product of the equation ( ) by $i\mathbf{k}$ gives

$$(\nu'|\mathbf{k}|^2-i\omega)i\mathbf{k}\cdot\widetilde{\mathbf{v}^*}-|\mathbf{k}|^2\cdot\frac{\widetilde{p^*}}{\rho_0}=\frac{1}{\rho_0}\int\int_{\mathcal{S}}i\mathbf{k}\cdot\mathbf{t}(\mathbf{x})\times\exp(-i\mathbf{k}\cdot\mathbf{x})da.\tag{A10}$$

The equations (A8), (A9), and (A10) can be solved for $\widetilde{p^*}$ and $\widetilde{\mathbf{v}^*}$,

$$\widetilde{p^*}(\mathbf{k})=\frac{1}{1-i\omega\nu'/c_0^2}\int\int_{\mathcal{S}}\frac{i\mathbf{k}\cdot\mathbf{t}(\mathbf{x})}{|\mathbf{k}|^2-k^2}\exp(-i\mathbf{k}\cdot\mathbf{x})da,\tag{A11}$$

$$\widetilde{\mathbf{v}^*}(\mathbf{k})=\frac{1}{\rho_0\nu}\int\int_{\mathcal{S}}\frac{\mathbf{t}(\mathbf{x})}{|\mathbf{k}|^2-k^{*2}}\exp(-i\mathbf{k}\cdot\mathbf{x})da\tag{A12}$$

$$+\frac{1}{i\omega\rho_0}\int\int_{\mathcal{S}}\frac{i\mathbf{k}[i\mathbf{k}\cdot\mathbf{t}(\mathbf{x})]}{|\mathbf{k}|^2-k^{*2}}\exp(-i\mathbf{k}\cdot\mathbf{x})da-\frac{1}{i\omega\rho_0}\int\int_{\mathcal{S}}\frac{i\mathbf{k}[i\mathbf{k}\cdot\mathbf{t}(\mathbf{x})]}{|\mathbf{k}|^2-k^2}\exp(-i\mathbf{k}\cdot\mathbf{x})da.\tag{A13}$$

For determining the corresponding representation formulas in physical space, we shall use the inversion formula[33,36]

$$\frac{1}{(2\pi)^3}\int\int\int\frac{\exp(i\mathbf{k}\cdot\mathbf{x})}{|\mathbf{k}|^2-\lambda}dv_k=\frac{\exp(i\sqrt{\lambda}|\mathbf{x}|)}{4\pi|\mathbf{x}|},\tag{A14}$$

where $\sqrt{\lambda}=\gamma_1+i\gamma_2$, $\gamma_2\geqslant 0$. There results

$$\gamma p(\mathbf{x}) = p^0(\mathbf{x}) - \boldsymbol{\nabla} \cdot \int \int_{\mathcal{S}} \frac{\mathbf{t}(\mathbf{x}')\exp(ik|\mathbf{x}-\mathbf{x}'|)}{4\pi(1-i\omega\nu'/c_0^2)|\mathbf{x}-\mathbf{x}'|} da', \tag{A15}$$

$$\gamma \mathbf{v}(\mathbf{x}) = \mathbf{v}^0(\mathbf{x}) + \int \int_{\mathcal{S}} \frac{\mathbf{t}(\mathbf{x}')\exp(ik^*|\mathbf{x}-\mathbf{x}'|)}{4\pi\rho_0\nu|\mathbf{x}-\mathbf{x}'|} da' + \boldsymbol{\nabla}\boldsymbol{\nabla} \cdot \int \int_{\mathcal{S}} \times \frac{\mathbf{t}(\mathbf{x})\,[\exp(ik^*|\mathbf{x}-\mathbf{x}'|) - \exp(ik|\mathbf{x}-\mathbf{x}'|)]}{4\pi i\omega\rho_0|\mathbf{x}-\mathbf{x}'|} da' \tag{A16}$$

The function $\gamma(\mathbf{x})$ appears due to the known property of Fourier recovery of discontinuous furnctions. It has the expression

$$\gamma(\mathbf{x}) = \begin{cases} 1, & \text{for } \mathbf{x} \in \mathcal{D}^-, \\ 0.5, & \text{for } \mathbf{x} \in \mathcal{S}, \\ 0, & \text{for } \mathbf{x} \in \mathcal{D}^+. \end{cases}$$

The relationship (A16) is the *fundamental integral formula* for the motion of a compressible viscous fluid in a linear acoustic approximation.

[1]T. Chapman and B. Webb, "A neuromorphic hair sensor model of wind-mediated escape in the cricket," Int. J. Neural Syst. **9**, 397–403 (1999).
[2]J. Tautz, "Reception of particle oscillation in a medium-an unortodox sensory capacity," Naturwiss. **66**, 452–461 (1979).
[3]J. Tautz and H. Markl, "Catepillars detect flying wasps by hairs sensitive to airborne vibration," Behav. Ecol. Sociobiol. **4**, 101–110 (1978).
[4]N. H. Fletcher, "Acoustical response of hair receptors in insects," J. Comp. Physiol. **127**, 185–189 (1978).
[5]A. Reisland and P. Gorner, "Trichobothria," in *Neurobiology of Arachnids*, edited by F. G. Barth (Springer-Verlag, Berlin, 1985), pp. 138–161.
[6]F. G. Barth, *A Spider's World. Senses and Behavior* (Springer-Verlag, Berlin, 2002), p. 394.
[7]C. E. Bond, *Biology of Fishes*, 2nd ed. (Saunders College Publishing, Philadelphia, 1996).
[8]J. Mogdans, J. Engelmann, W. Hanke, and S. Krother, "The fish lateral line: How to detect hydrodynamic stimuli," in *Sensors and Sensing in Biology and Engineering*, edited by F. G. Barth, J. A. C. Humphrey, and T. W. Secomb (Springer-Verlag, Vienna, 2003), pp. 173–185.
[9]T. Shimozawa and M. Kanou, "Varieties of filiform hairs: range fractionation by sensory afferents and cercal interneurons of the cricket *Grillus Bimaculatus*," J. Comp. Physiol., A **155**, 495–505 (1984).
[10]I. Imai, "A new method of solving Oseen's equations and its application to the flow past an inclined elliptic cylinder," Proc. R. Soc. London, Ser. A **224**, 141–160 (1954).
[11]T. Kumagai, T. Shimozawa, and Y. Baba, "The shape of wind receptor hairs of cricket and cockroach," J. Comp. Physiol., A **183**, 187–192 (1998).
[12]J. A. C. Humphrey, R. Devarakonda, I. Iglesias, and F. G. Barth, "Dynamics of arthropod filiform hairs. I. Mathematical modeling of the hair and air motion," Philos. Trans. R. Soc. London, Ser. B **340**, 423–444 (1993).
[13]G. G. Stokes, "On the effect of the internal friction of fluids on the motion of pendulums," Trans Camb Phil Soc 9, 8ff (1851) [*reprinted in Mathematical and Physical Papers*, (Cambridge University Press, Cambridge, 1901), Vol. **III**, pp. 1–141.
[14]J. A. C. Humphrey, R. Devarakonda, I. Iglesias, and F. G. Barth, "Errata for dynamics of arthropod filiform hairs. I. Mathematical modeling of the hair and air motion," Philos. Trans. R. Soc. London, Ser. B **352**, 1995 (1997).
[15]J. A. C. Humphrey, R. Devarakonda, I. Iglesias, and F. G. Barth, "Errata for dynamics of arthropod filiform hairs. I. Mathematical modeling of the hair and air motion," Philos. Trans. R. Soc. London, Ser. B **353**, 2163 (1998).
[16]F. G. Barth, U. Wastl, J. A. C. Humphrey, and R. Devarakonda, "Dynamics of arthropod filiform hairs. III. Mechanical properties of spider trchobothria (*Cupiennius salei Keis.*)," Philos. Trans. R. Soc. London, Ser. B **340**, 445–461 (1993).
[17]F. G. Barth, J. A. C. Humphrey, J. Halbritter, and W. Brittinger, "Dynamics of arthropod filiform hairs. II. Flow patterns related to air movement detection in spider (*Cupiennius salei Keis.*)," Philos. Trans. R. Soc. London, Ser. B **347**, 397–412 (1995).
[18]R. Devarakonda, F. G. Barth, and J. A. C. Humphrey, "Dynamics of arthropod filiform hairs. IV. Hair motion in air and water," Philos. Trans. R. Soc. London, Ser. B **351**, 933–946 (1996).
[19]F. G. Barth and A. Holler, "Dynamics of arthropod filiform hairs. V. The response of spider tricobothria to natural stimuli," Philos. Trans. R. Soc. London, Ser. B **354**, 183–192 (1999).
[20]T. Shimozawa, T. Kumagai, and Y. Baba, "Structural scaling and function design of the cercal windreceptor hairs of cricket," J. Comp. Physiol., A **183**, 171–186 (1998).
[21]J. A. C. Humphrey, F. G. Barth, and K. Voss, "The motion-sensing hairs of Arthropods: using physics to understand sensory ecology and adaptive evolution," in *Ecology of Sensing*, edited by F. G. Barth and A. Schmid (Springer-Verlag, Berlin, Heidelberg 2001).
[22]J. A. C. Humphrey, F. G. Barth, M. Reed, and A. Spak, "The physics of Arthropod medium-flow sensitive hairs: biological models for artificial sensors," in *Sensors and Sensing in Biology and Engineering*, edited by F. G. Barth, J. A. C. Humphrey, and T. W. Secomb, (Springer-Verlag, Berlin 2003), pp. 129–144.
[23]T. Shimozawa, J. Murakami, and T. Kumagai, "Cricket wind receptors: thermal noise for the highest sensitivity known," in *Sensors and Sensing in Biology and Engineering*, edited by F. G. Barth, J. A. C. Humphrey, and T. W. Secomb (Springer-Verlag, Berlin, 2003).
[24]F. G. Barth and H-E Dechant, "Arthropod cuticular hairs: tactile sensors and refinement of stimulus transformation," 159–171.
[25]Y. Ozaki, T. Ohyama, T. Yasuda, and I. Shimoyama, "An air flow sensor modeled on wind receptor hairs of insects," *Proc. MEMS* 2000, Miazaki, Japan, pp. 531–537.
[26]J. Chen, Z. Fan, J. Engel, and C. Liu, "Two-dimensional micromachined flow sensor array for fluid mechanics studies," J. Aerosp. Eng. **16**, 85–97 (2003).
[27]Z. Fan, J. Chen, J. Zou, D. Bullen, C. Liu, and F. Delcomyn, "Design and fabrication of artificial lateral line flow sensors," J. Micromech. Microeng. **12**, 665–661 (2002).
[28]J. van Baar, M. Dijkstra, R. Wiegerink, T. Lammerink, R. de Boer, and G. Krijnen, "Artificial sensory hairs based on the flow sensitive receptor hairs of crickets," J. Micromech. Microeng. **15**, 132–138 (2005).
[29]J. Zou, J. Chen, and C. Liu, "Plastic deformation magnetic assembly (PDMA) of out-of-plane microstructures: technology and application," J. Microelectromech. Syst. **10**, 302–309 (2001).
[30]O. Rudko, M.Sc. thesis, State University of New York at Binghamton, "Design of a low-frequency sound sensor inspired by insect sensory hairs," Department of Mechanical Engineering, 2001.
[31]D. Homentcovschi and R. N. Miles, "Viscous scattering of a pressure wave by a hard body," submitted for publication.
[32]L. Dragos, *Mathematical Methods in Aerodynamics* (Springer-Verlag, Berlin, 2004).
[33]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, New Jersey, 1968).
[34]A. D. Pierce, *Acoustics* (McGraw-Hill, New York, 1981).
[35]L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, New York, 1959).
[36]I. Stakgold, *Boundary Value Problems of Mathematical Physics* (SIAM, Philadelphia, 2000).
[37]C. Pozrikidis, *Boundary Integral and Singularity Methods for Linearized Viscous Flow* (Cambridge University Press, Cambridge, 1992).
[38]W. Olmstead and A. K. Gautesen, "A new paradox in viscous hydrodynamics," Arch. Ration. Mech. Anal. **29**, 58–65 (1968).

Modal density of rectangular volumes, areas, and lines

R. D. Blevins[a)]
Goodrich, San Diego, California, 91910

(Received 24 May 2005; revised 25 September 2005; accepted 27 November 2005)

Exact solution for the cumulative number of modes in a rectangular volume with reflective walls is compared with asymptotic estimates and limit cases. At low frequencies, the asymptotic expressions do not provide the correct limits and deviate appreciably from the exact results and the exact solution offers a computationally efficient alternative. The results are extended to two-dimensional regimes, membranes, plates, beams, and strings.
 [DOI: 10.1121/1.2159293]

PACS number(s): 43.20.Ks [ADP] Pages: 788–791

I. INTRODUCTION

Consider the closed rectangular volume filled with a homogeneous inviscid compressible fluid shown in Fig. 1(a). Fluid pressure (p) modes satisfy the modal form of the acoustic wave equation, $\nabla^2 p+(\omega/c)^2 p=0$. There are two sets of boundary conditions: (1) open boundary, $p=0$ on the sides, which is called the Dirichlet problem, and (2) closed (rigid) boundaries, where $\partial \mathbf{p}/\partial \mathbf{n}=0$ on the sides where $\mathbf{n}$ is the outward normal to the side, which is called the Neumann problem. Baltes and Hilf (1976)—also see Maa (1939), Bolt (1939), Morse and Ingard (1968), and Pierce (1981)—give asymptotic formulas as frequency f (Hertz) becomes large for the cumulative number of acoustic modes N in a rectangular fluid volume below frequency f,

$$N^{3D}_{\text{asymptotic}} = \begin{cases} \dfrac{4\pi f^3 V}{3c^3} - \dfrac{\pi f^2 S}{4c^2} + \dfrac{fL}{8c} & \text{open boundary} \quad (1a) \\ \dfrac{4\pi f^3 V}{3c^3} + \dfrac{\pi f^2 S}{4c^2} + \dfrac{fL}{8c} & \text{closed boundary,} \quad (1b) \end{cases}$$

where c is speed of sound of fluid, f the maximum acoustic frequency in Hertz, L the sum of the lengths of all straight edges for a rectangular volume, $4(L_x+L_y+L_z)$, S the total surface area of the rectangular volume, $2(L_xL_y+L_yL_z+L_xL_z)$, V the rectangular volume, $L_xL_yL_z$, and N is the number of acoustic modes with natural frequency below f.

For two-dimensional analyses, the fluid is contained between two closely spaced parallel, rigid rectangular sheets and the boundary conditions are applied to the exposed fluid edge (Baltes and Hilf, 1976),

$$N^{2D}_{\text{asymptotic}} = \begin{cases} \pi A f^2/c^2 - Pf/(2c) & \text{open edges} \quad (2a) \\ \pi A f^2/c^2 + Pf/(2c) & \text{closed edges.} \quad (2b) \end{cases}$$

Here, A is the area of a two-dimensional area, L_xL_y. L_x, L_y is the length and breadth of the two-dimensional rectangle shown in Fig. 1(b). P is the perimeter of two-dimensional area, $2(L_x+L_y)$.

The purpose of this paper is to explore the validity of these equations in comparison to exact, asymptotic, and limit cases.

[a)]Electronic mail: rdblevins@aol.com

II. LIMIT CASES

Courant and Hilbert (1924, 1953) showed that the asymptotic limit cases for rectangular volume and rectangular area as frequency approaches infinity are given by

$$\lim_{f\to\infty} N = \begin{cases} 4\pi V f^3/3c^2 & \text{3D} \quad (3a) \\ \pi A f^2/c^2 & \text{2D.} \quad (3b) \end{cases}$$

These are the highest order frequency terms in Eqs. (1) and (2) and they dominate at high frequencies. At lower frequencies, the lower order terms dominate Eqs. (1) and (2).

Consider the case of a long, slender, closed rectangular duct whose width L_z and height L_y are small compared to its length, L_x, $L_z=L_y=\delta \ll L_x$, shown in Fig. 1(c),

$$\begin{aligned} L &= 4(L_x+L_y+L_z) = 4L_x + 8\delta, \\ S &= 2(L_xL_y + L_xL_z + L_yL_z) = 4L_x\delta + 2\delta^2, \\ V &= L_xL_yL_z = L_x\delta^2. \end{aligned} \quad (4)$$

In the limit of a slender duct as L_x/δ approaches infinity, only the edge length in Eq. (1) remains,

$$\lim_{\delta\to 0} V = 0, \quad \lim_{\delta\to 0} S = 0, \quad \lim_{\delta\to 0} L = 4L_x, \quad (5)$$

and Eq. (1) reduces to the edge term for both open and closed boundaries,

$$N^{\text{3D-asymptotic}}_{L_y=L_z=0} = Lf/8c = fL_x/2c. \quad (6)$$

This can be compared with the exact result. The natural frequencies of acoustic modes in a long, narrow closed duct are obtained by setting the indices j and k to zero in Eq. (9b), to obtain natural frequencies $f_i=ic/(2L_x)$, $i=0,1,2,3,\ldots$ equally spaced at intervals of $\Delta f=c/(2L_x)$. The total number of duct modes below a given frequency f is $f/\Delta f$,

$$N_{\text{1D-exact}} = 2L_x f/c. \quad (7)$$

Equations (6) and (7) differ by a factor of 4. Since Eq. (6) is exact, Eq. (1) is incorrect for one-dimensional acoustic modes. Equation (2) reduces to the one-dimensional modes when L_y is set to zero,

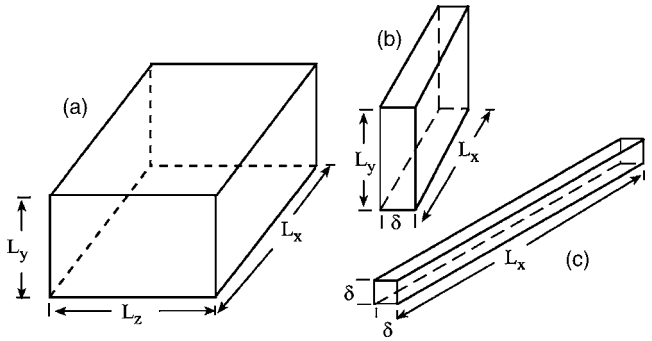


FIG. 1. Rectangular acoustic volumes.

$$N^{\text{2D-asymptotic}}_{L_y=0,\ \text{closed sides}} = \begin{cases} -L_x f/c & \text{open edges} \\ L_x f/c & \text{closed edges.} \end{cases} \tag{8}$$

These results do not match the exact result, which is 0 for open edges and $2L_x f/c$ for closed edges. The middle terms of Eq. (1) also do not successfully reduce to their two-dimensional limit cases. Thus Eq. (1) may be less useful at low frequency, where one- or two-dimensional modes dominate, than at high frequency, where three-dimensional modes dominate.

III. EXACT SOLUTIONS

The acoustic natural frequencies in Hertz of the three-dimensional rectangular volume of Fig. 1(a) are given by Rayleigh (1896),

$$f_{ijk} = \frac{c}{2}\left(\frac{i^2}{L_x^2} + \frac{j^2}{L_y^2} + \frac{k^2}{L_z^2}\right)^{1/2} \tag{9}$$

$i = 1,2,3, \ldots, \quad j = 1,2,3, \ldots, \quad k = 1,2,3, \ldots,$

open boundary

$i = 0,1,2,3, \ldots, \quad j = 0,1,2,3, \ldots, \quad k = 0,1,2,3, \ldots,$

closed boundary

Each integer index triplet (i,j,k) defines a unique acoustic mode. The associated acoustic natural frequencies, f_{ijk}, may have repeated values. The closed volume supports nontrivial modes with one or two zero indices that have constant pressure along the corresponding orthogonal axes. The open boundary does not support one- or two-dimensional modes.

By sorting the edge lengths so that $L_x \geqslant L_y \geqslant L_z$, Eq. (9) defines the ascending frequencies that scale the cumulative acoustic modal density of a closed rectangular volume,

$$\text{null mode} \quad f_{000} = 0, \tag{10a}$$

$$\text{first edge mode} \quad f_{100} = c/(2L_x), \tag{10b}$$

$$\text{first side mode} \quad f_{110} = (c/2)(1/L_x^2 + 1/L_y^2)^{1/2}, \tag{10c}$$

$$\text{first oblique mode} \quad f_{111} = (c/2)(1/L_x^2 + 1/L_y^2 + 1/L_z^2)^{1/2}. \tag{10d}$$

There are no modes between the null mode and the first edge mode, f_1. There are only one-dimensional edge modes at low frequency, below f_{110}. Side modes dominate between f_{110} and f_{111}. Interior modes dominate at high frequency, above f_{111}.

This index ellipsoid is defined by setting $f \geqslant f_{ijk}$ in Eq. (9), dividing through by f, and then squaring both sides, Courant and Hilbert (1924, 1953),

$$1 \geqslant \frac{i^2}{i_{\max}^2} + \frac{j^2}{j_{\max}^2} + \frac{k^2}{k_{\max}^2}, \tag{11}$$

$$i = 0,1,2, \ldots, I, \quad j = 0,1,2, \ldots, J, \quad k = 0,1,2, \ldots, K,$$

$$i_{\max} = 2L_x f/c, \quad j_{\max} = 2L_y f/c, \quad k_{\max} = 2L_z f/c. \tag{12}$$

$i_{\max}$, $j_{\max}$, and $k_{\max}$ give the acoustic half wavelengths per edge; Their integer values I,J,K are the number of modes along each index axis. The maximum value of one index given the other two are found by solving Eq. (11).

$$I(j,k) = \text{Int}[i_{\max}(1 - j^2/j_{\max}^2 - k^2/k_{\max}^2)^{1/2}],$$

$$J(i,k) = \text{Int}[j_{\max}(1 - i^2/i_{\max}^2 - k^2/k_{\max}^2)^{1/2}],$$

$$K(i,j) = \text{Int}[k_{\max}(1 - i^2/i_{\max}^2 - j^2/j_{\max}^2)^{1/2}], \tag{13}$$

$$I = I(0,0) = \text{Int}[i_{\max}], \quad J = J(0,0) = \text{Int}[j_{\max}],$$

$$K = K(0,0) = \text{Int}[k_{\max}].$$

The integer function Int( ) rounds down to the nearest integer. For example, $\text{Int}(\pi) = 3$.

The total number of acoustic modes with natural frequencies equal to or less than f in a rectangular volume is a triple summation of modes over the three modal indices,

$$N(f) = \sum_{k=0}^{K}\sum_{j=0}^{J}\sum_{i=0}^{I} \delta_{ijk} \quad \text{where } \delta_{ijk} = \begin{cases} 1 & \text{if } f \geqslant f_{ijk} \\ 0 & \text{if } f < f_{ijk}. \end{cases} \tag{14}$$

N is an integer and it includes the null mode, $f_{000} = 0$. Equation (14) can be reordered to eliminate the zero frequency null modes. Equation (13) then replaces the innermost summations, reducing the order of the number of computations from order IJK to KJ, which is most efficient if $L_x > L_y > L_z$,

$$N^{\text{3D-exact}}_{\text{open}}(f) = N_{\text{oblique}} = \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{I} \delta_{ijk} = \sum_{k=1}^{K}\sum_{j=1}^{J(0,k)}\sum_{i=1}^{I(j,k)} 1 = \sum_{k=1}^{K}\sum_{j=1}^{J(0,k)} I(j,k), \tag{15a}$$

$$N^{\text{3D-exact}}_{\text{closed}}(f) = N_{\text{oblique}} + N_{\text{sides}} + N_{\text{edges}} = \sum_{k=0}^{K}\sum_{j=0}^{J(0,k)}\sum_{i=0}^{I(j,k)} 1 = \sum_{k=0}^{K}\sum_{j=0}^{J(0,k)} (I(j,k) + 1), \tag{15b}$$

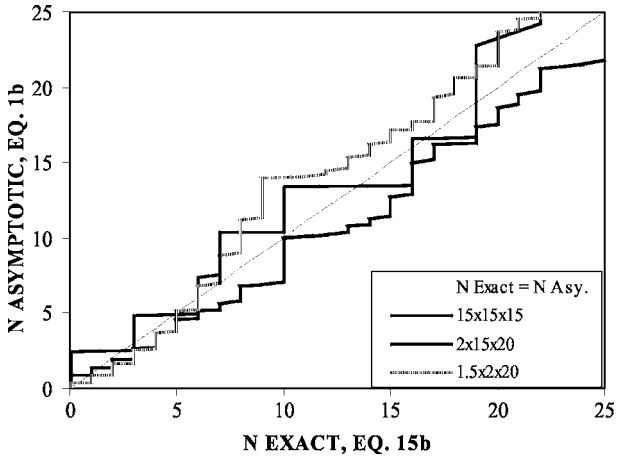


FIG. 2. Exact and asymptotic solutions for cumulative modal density of the first 25 modes of $15\times15\times15$ ft, $2\times15\times20$ ft, and $1.5\times2\times20$ ft closed rooms. 1 m=3.28 ft. c=1128 ft/s (343 m/s).

$$N_{\text{sides-exact}} = \sum_{j=1}^{J}\sum_{i=1}^{I}\delta_{ij0} + \sum_{k=1}^{K}\sum_{i=1}^{I}\delta_{i0k} + \sum_{k=1}^{K}\sum_{j=1}^{J}\delta_{0jk} = \sum_{j=1}^{J} I(j,0) + \sum_{k=1}^{K} I(0,k) + \sum_{k=1}^{K} J(0,k), \tag{15c}$$

$$N_{\text{edges-exact}} = \sum_{i=1}^{I}\delta_{i00} + \sum_{j=1}^{J}\delta_{0j0} + \sum_{k=1}^{K}\delta_{00k} = I + J + K. \tag{15d}$$

Oblique refers to three-dimensional interior modes with i, j, and k greater than zero and "Side" refers to two-dimensional modes with only one of i, j, or k equal to zero. "Edge" refers to one-dimensional axial modes with two of i, j, and k equal to zero. These lie on the index axes.

Equation (15b) is compared with Eq. (1b) in Fig. 2 for the first 25 modes of three closed acoustic volumes, neglecting the null mode. There is appreciable difference between the exact and asymptotic results. Figure 3 shows two cases from Fig. 2 over a broader frequency range which has been nondimensionalized by the first oblique mode, Eq. (10d). The exact solution, Eq. (15b), is compared with the asymptotic estimate, Eq. (1b), and Courants limit case (3a). The asymptotic solution provides an increasingly good estimate of the exact result as frequency increases beyond the frequency of the first three-dimensional interior oblique mode f_{111}. It approaches the exact result for frequencies on the order of 10 f_{111} and greater. But below this frequency there are significant deviations. Courant's asymptotic solution [Eq. (3a)] significantly underestimates the cumulative modal density of closed three-dimensional rectangular volumes for frequencies on order of f_{111} and lower.

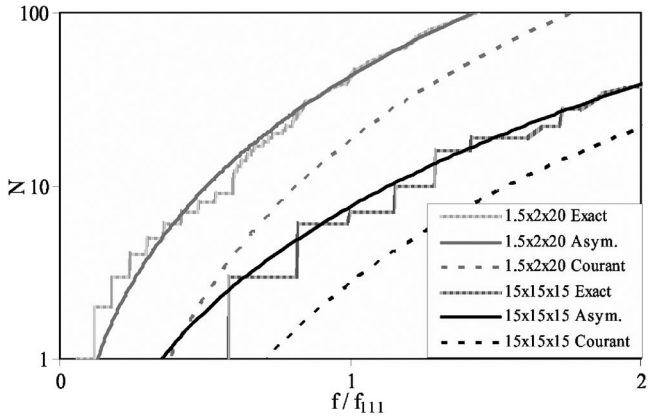


FIG. 3. Cumulative modal density for two closed rectangular volumes as a function of frequency nondimensionalized by f_{111} [Eq. (10d)]. Equations (1b), (3a), and (15b).

The exact solution for closed boundaries can be reduced to two dimensions by setting L_z=0 in Eq. (15b),

$$N^{\text{2D-exact}}_{\text{open edge}}(f) = \sum_{j=1}^{J}\sum_{i=1}^{I(j)}\delta_{ij0} = \sum_{j=1}^{J} I(j,0) = \sum_{j=1}^{\text{Int}[2L_y f/c]} \text{Int}[(2L_x f/c)(1 - j^2/(2L_y f/c)^2)^{1/2}],$$

$$N^{\text{2D-exact}}_{\text{closed edge}}(f) = \sum_{j=0}^{J}\sum_{i=0}^{I(j)}\delta_{ij0} = \sum_{i=0}^{J} I(j,0) + I + J = \sum_{j=0}^{\text{Int}[2L_y f/c]} \text{Int}[(2L_x f/c)(1 - j^2/(2L_x f/c)^2)^{1/2} + 1]. \tag{16}$$

The open boundary is only applied at the edges. The closed edge number of modes equals the modes of a single side given in Eq. (15b). Certain symmetries exist, for example $\Sigma_{j=1}^{J} I(j,0) = \Sigma_{i=1}^{I} J(i,0)$. In one-dimensional duct analysis, the fluid is contained in a constant area, rigid-walled duct of length L; the open or closed boundaries are applied at the two extreme ends,

$$N_{\text{1D-exact}}(f) = \sum_{i=0}^{I}\delta_{i00} = I = \text{Int}[2Lf/c] \quad \text{both ends open or closed.} \tag{17}$$

This result is obtained by setting L_y=L_z=0 in Eqs. (15b) and (15d) or (16b).

IV. MEMBRANES, PLATES, BEAMS, AND STRINGS

The wave equation describes the propagation of acoustic waves, out-of plane membrane waves, and waves in strings. The open boundary condition for acoustics is equivalent to a supported edge for membranes and the closed boundary condition is equivalent to a guided (zero slope) edge. Equation (16) apples to flat rectangular membranes with the substitution of the membrane wave speed and Eq. (17) applies to strings with the substitution of the string wave speed.

The expression for the natural frequency in Hertz of out-of-plane bending modes of simply supported rectangular plate where E is the elastic modulus, γ is the mass per unit area, and ν is Poisson's ratio (Blevins, 1979),

TABLE I. Exact and asymptotic estimates for cumulative acoustic modal density of $10\times15\times30$ ft closed room. N_{exact} is Eq. (15b).

Freq (Hz)	10	18.80	42.04	50	70.34	200	500	703.4	1000	2000	5000	10000
f/f_{111}	0.142	0.2673	0.598	0.711	1	2.843	7.108	10	14.22	28.43	71.08	142.16
N-exact	0	1	4	4	12	157	1922	5141	14 272	109 571	1 669 665	13 245 011
N-Eq. (1b)	0.368	0.938	3.964	5.638	11.78	154.39	1931.6	5137.6	14 268	109 559	1 669 561	13 244 650
N-Eq. (3b)	0.013	0.0873	0.976	1.642	4.571	105.07	1641.6	4570.6	13 133	105 066	1 641 662	13 133 299

$$f_{ij}=\frac{\pi}{2}\left(\frac{i^2}{L_x^2}+\frac{j^2}{L_y^2}\right)\sqrt{\frac{Eh^3}{12\gamma(1-\nu^2)}},$$
$$i=1,2,3,\ldots,\quad j=1,2,3,\ldots, \tag{18}$$

can be put in the same form as Eq. (11) if f_{ij} is replaced by its upper bound $f\leqslant f_{ij}$, the equation is divided through by f, and $i_{\max}$ and $j_{\max}$ are redefined in terms of the plate

$$1\geqslant\frac{i^2}{i_{\max}^2}+\frac{j^2}{j_{\max}^2},\quad i=1,2,3,\ldots I,\quad j=1,2,3,\ldots,J,$$
$$i_{\max}=\frac{2L_xf^{1/2}}{c_e},\quad j_{\max}=\frac{2L_yf^{1/2}}{c_e}, \tag{19}$$
$$c_e=\left[2\pi\left(\frac{Eh^3}{12\gamma(1-\nu^2)}\right)^{1/2}\right]^{1/2},\quad c=c_ef^{1/2}.$$

The previous solutions, Eqs. (2a) and (16a), can then be applied,

$$N_{\substack{\text{simply}\\\text{supported}\\\text{plate}}}=\begin{cases}\pi L_xL_yf/c_e^2-(L_x+L_y)f^{1/2}/c_e\\ \text{asymptotic}\\ \displaystyle\sum_{j=1}^{\text{Int}[2f^{1/2}L_y/c_e]}\text{Int}\left[\frac{2L_xf^{1/2}}{c_e}\left(1-j^2\left(\frac{c_e}{2L_yf^{1/2}}\right)^2\right)^{1/2}\right]\\ \text{exact.}\end{cases} \tag{20}$$

The modulus of elasticity of the plate material is E and ν is Poisson's ratio. The speed of bending wave propagation in thin plates is proportional to the square root of frequency, $c_{\text{plate}}=c_ef^{1/2}$ where c_e has units of length/time$^{1/2}$. The modal density, dN/df, of plates does not increase with frequency as do two-dimensional acoustic modes or membrane modes.

The duct analysis of Eq. (17) can be used to predict the cumulative modal density of bending modes of a pined-pined, slender, elastic beam with length L by substituting the beam bending wave speed of the beam, $c=c_ef^{1/2}$ where $c_e=(2\pi)^{1/2}[EI_c/m]^{1/4}$, into Eq. (17).

$$N_{\text{exact}}^{\text{beam}}=\text{Int}[Lf^{1/2}/((\pi/2)^{1/2}(EI_c/m)^{1/4})]. \tag{21}$$

I_c is the area moment of inertia of the cross section and m is mass per unit length.

V. DISCUSSION AND CONCLUSION

The terms on the right-hand side of Eq. (1), beyond Courant's volume term [Eq. (3)], are essentially corrections to Courant's solution to improve Eq. (1)'s predictions at low frequency. While producing a significant improvement in prediction below 10 f_{111}, these correction terms are not robust in that they do not produce the correct limit for one- or two-dimensional modes and the results can deviate significantly from the exact result at and below the first oblique mode f_{111} [Eq. (10d)].

The computation effort to exactly compute the cumulative modal density below 10 f_{111} is relatively modest. This can be seen for the $10\times15\times30$ ft closed room (Morse and Ingard, 1968), in Table I. This particular geometry has a very favorable comparison with the asymptotic solution [Eq. (1b)]. The lengths are ordered, $L_x=30$ ft$>L_y=15$ ft$>L_z=10$ ft, so $f_{100}=18.80$ Hz, $f_{110}=42.04$ Hz, and $f_{111}=70.34$ Hz. At 10 $f_{111}=703.4$ Hz, $I=37$, $J=18$, $K=12$. $JK=216$ terms are evaluated to compute the exact cumulate modal density at 10 f_{111}.

As frequencies increase above 10 f_{111}, the asymptotic expressions of Courant [Eq. (3)] or Eq. (1) provide an accurate estimate.

ACKNOWLEDGMENTS

This paper benefited greatly from the advice of Dr. H. W. Kwan, Mary Carlsson, Boecky Yalof, and the reviewers.

Baltes, H. P., and Hilf, E. R. (**1976**). *Spectra of Finite Systems* (Bibliographisches Institut Mannhein/Wein/Zurich), ISBN 3-411-01491-1. pp. 51, 59.

Blevins, R. D. (**1979**). *Formulas for Natural Frequency and Mode Shape* (reprinted by Kreiger).

Bolt, R. H. (**1939**). "Frequency distribution of eigentones in a three-dimensional continuum," J. Acoust. Soc. Am. **16**, 228–234.

Courant, R., and Hilbert, D. (**1924**). *Methoden der Mathematischen Physik*, 1st German ed. (Springer, Berlin), pp. 344–346.

Courant, R., and Hilbert, D. (**1953**). *Methods of Mathematical Physics*, First English ed. (Interscience, New York), Vol. **I**, pp. 430–442.

Maa, D.-Y. (**1939**). "Distribution of eigentones in a rectangular chamber at low frequency range," J. Acoust. Soc. Am. **16**, 235–238.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), pp. 582–588.

Pierce, A. D. (**1981**). *Acoustics, An Introduction to its Physical Principals and Applications* (McGraw-Hill, New York), p. 293.

Rayleigh, J. W. S. (**1896**). *The Theory of Sound* (Dover, New York, **1945**, reprint of 2nd edition of 1896), Vol. **II**, p. 70.

Probe-corrected near-field measurements on a truncated cylinder[a)]

Thorkild B. Hansen[b)]
Consultant, Air Force Research Laboratory (SNHE), Hanscom AFB, Massachusetts 01731

(Received 3 March 2005; revised 17 November 2005; accepted 18 November 2005)

A probe-corrected theory is presented for computing the acoustic far fields of transducers and scatterers from measurements of near fields on a cylindrical surface. The near-field data is truncated at the top, bottom, and angular edges of the scan cylinder. These truncation edges can cause inaccuracies in the computed far fields. Correction techniques are developed for the top and bottom truncation edges. The cylindrical wave expansions automatically apply an angular taper to the near-field data that reduces the effect of the angular truncation edges. The taper function depends on the probe and the angular sample spacing. The theory is validated through numerical examples involving a point source and a baffled piston transducer probe. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151789]

PACS number(s): 43.20.Rz, 43.20.Bi, 43.20.Ye [GCG] Pages: 792–807

I. INTRODUCTION

Probe-corrected near-field techniques have been widely used for the past 40 years to characterize antennas and transducers from measurements on planar,[1–5] cylindrical,[6–10] and spherical[11–20] scanning surfaces. The field of the antenna or transducer is first measured with a known probe on the scanning surface in the near field. Probe-corrected formulas are then applied to the measured near field to get the desired far field of the antenna or transducer. The measurements are typically performed in anechoic chambers. The electromagnetic probe-corrected formulas have been derived and widely implemented for scanning on planar, cylindrical, and spherical surfaces. The corresponding acoustic formulas have only been derived for scanning on planar and spherical surfaces.

The far field of electromagnetic and acoustic scatterers can also be determined from near-field measurements in anechoic chambers.[21–23] A far-field source or a compact-range reflector can provide the plane-wave field required to illuminate or insonify the scatterers. To compute the scattered far field, the incident field and background interactions must be extracted from the measured total field. The simplest method for eliminating these field contributions is standard background subtraction which involves two measurements. The first measurement is carried out with the scatterer present, and the second measurement is carried out with the scatterer removed. The results of the two measurements are then subtracted to get an approximation for the scattered near field, which can be processed with the probe-corrected formulas to get the scattered far field.

It is appropriate to mention the related area of research known as near-field acoustical holography, where near-field measurements are employed to backpropagate the pressure field in space and time toward the source.[24–27] One can achieve super-resolution with near-field holography when part of the evanescent spectrum is captured during the scan. In this paper, however, we are concerned only with computing the field of the source outside the scanning surface and do not attempt to backpropagate the near-field data.

In this paper we derive the probe-corrected formulas for cylindrical near-field scanning that can be used to compute acoustic far fields of transducers and scatterers. We present computation schemes and sampling theorems that allow the far fields to be computed from sampled values of the probe output on the scan cylinder. In practice, the measurements are carried with a mechanical scanner over a truncated scan cylinder that covers only a limited region of the infinite scan cylinder that is used for the derivation. Therefore we carefully examine the probe-corrected formulas to determine how truncation edges may affect the far-field accuracy. We develop edge-correction techniques that reduce the far-field errors caused by truncation and show that an angular taper function is automatically applied to the probe output. The results of the paper are validated through numerical examples involving a point source and a baffled piston transducer probe with a circular cross section.

The paper is organized as follows. In Sec. II we derive the formulas for cylindrical near-field scanning of acoustic fields with an arbitrary known probe. These formulas express the field of the source region in terms of the probe receiving coefficients and the probe output over an infinite scan cylinder. For use in the numerical simulations, we derive exact expressions for the probe output when the probe is a baffled circular receiving piston transducer.

We consider a truncated scan cylinder in Sec. III and develop edge-corrected formulas that compensate for the end-point contributions to the far field from the top and bottom edges. We further study the angular taper function that is inherent in the cylindrical expansions. In Sec. IV the far-field formulas are converted into discrete computation schemes. In Sec. V we consider a point source at the origin to illustrate the implications of truncating the scan cylinder. Section VI presents conclusions.

[a)]Revised version submitted to the Journal of the Acoustical Society of America October 2005. Original manuscript submitted February 2005.
[b)]Electronic mail: Thorkild.Hansen@att.net

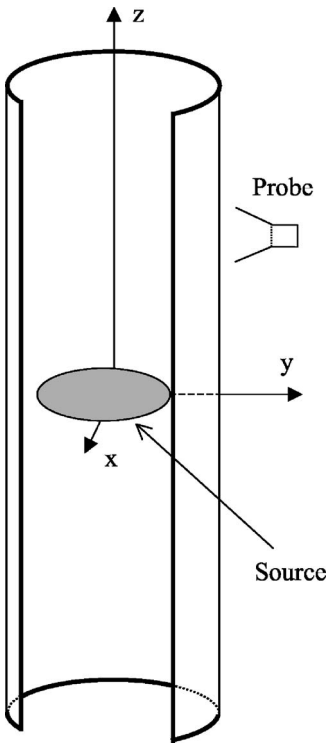


FIG. 1. Probe measures the field of the source on the truncated scan cylinder given by $\rho=a$, $-z_0<z<z_0$, $\phi_0<\phi<2\pi-\phi_0$.

II. THEORY OF CYLINDRICAL NEAR-FIELD SCANNING

We now derive the formulas for computing the acoustic field of a source of finite extent from either the pressure field (Sec. II A) or from the output of a known probe (Sec. II B) on a scan cylinder. The source can either be a transducer or a scatterer. For scatterers we assume that background subtraction has been performed, so that the incident field and background reflections have been removed. The part of space not occupied by the sources is a linear lossless fluid in which the acoustic field can be fully described by its pressure field $p(\mathbf{r})$.

As shown in Fig. 1, the field is measured on the cylinder $\rho=a$, where ρ is the cylindrical coordinate that equals the distance to the z axis. The scan cylinder in Fig. 1 is truncated and covers the region $-z_0\leqslant z\leqslant z_0$, $\phi_0\leqslant\phi\leqslant 2\pi-\phi_0$. For the derivations of this section the scan cylinder is assumed infinite ($z_0=+\infty$ and $\phi_0=0$).

We define the minimum cylinder $\rho=R_{\min}$ such that the maximum (supremum) value of the coordinate ρ for all points on the source equals $R_{\min}$. Note that $R_{\min}$ depends on the size of the source as well as on its location. For example, the value of $R_{\min}$ is equal to $\sqrt{x_1^2+y_1^2}$ for a point source at $(x,y,z)=(x_1,y_1,z_1)$. We assume that the scan cylinder encloses the minimum cylinder.

A. The far field in terms of the field on the scan cylinder

The formulas that express the field from a confined source in terms of the pressure on an enclosing cylinder will be derived in this section from the standard cylindrical wave expansion. With $e^{-i\omega t}$ time dependence suppressed, the basic outgoing solution $p^{(3)}(\mathbf{r})$ to the scalar wave equation

$$(\nabla^2+k^2)p(\mathbf{r})=0 \tag{1}$$

is

$$p^{(3)}(\rho,\phi,z)=H_n^{(1)}(k_\rho\rho)e^{in\phi}e^{ik_z z}, \tag{2}$$

where k is the wave number, k_z is a real parameter, n is an integer, and

$$k_\rho=\sqrt{k^2-k_z^2} \tag{3}$$

is a complex parameter with non-negative real and imaginary parts. Moreover, $H_n^{(1)}(k_\rho\rho)$ is the Hankel function of the first kind and order n, and (ρ,ϕ,z) are the standard cylindrical coordinates given in terms of rectangular coordinates (x,y,z) as

$$x=\rho\cos\phi,\quad y=\rho\sin\phi,\quad z=z. \tag{4}$$

Outside the minimum cylinder of radius $R_{\min}$, the field $p(\mathbf{r})$ of any source of finite extent can be expressed as a superposition of the basis fields in (2) as[28]

$$p(\rho,\phi,z)=\sum_{n=-\infty}^{+\infty}e^{in\phi}\int_{-\infty}^{+\infty}F_n(k_z)H_n^{(1)}(k_\rho\rho)e^{ik_z z}dk_z,$$

$$\rho>R_{\min}, \tag{5}$$

where $F_n(k_z)$ is the spectrum that characterizes the source.

The formula (5) determines the field everywhere outside a cylinder of radius $R_{\min}$ in terms of an integration and a summation involving the spectrum $F_n(k_z)$. By applying the method of stationary phase,[29] one can show that for observation points far form the source the field can be expressed in spherical coordinates as

$$p(r,\theta,\phi)\sim\frac{2e^{ikr}}{r}\sum_{n=-\infty}^{+\infty}F_n(k\cos\theta)e^{in\phi}e^{-i\pi(n+1)/2}, \tag{6}$$

where (r,θ,ϕ) are the standard spherical coordinates related to the rectangular coordinates (x,y,z) through

$$x=r\sin\theta\cos\phi,\quad y=r\sin\theta\sin\phi,\quad z=r\cos\theta. \tag{7}$$

Unlike the general formula (5), the far-field formula (6) does not involve an integration over k_z.

An expression for the spectrum $F_n(k_z)$ in terms of the field $p(a,\phi,z)$ on the scan cylinder is obtained by multiplying Eq. (5) by $e^{-in'\phi}e^{-ik_z' z}$ and integrating over the scan cylinder

$$F_n(k_z)=\frac{1}{4\pi^2H_n^{(1)}(k_\rho a)}\int_{-\infty}^{+\infty}\int_0^{2\pi}p(a,\phi,z)e^{-in\phi}e^{-ik_z z}d\phi dz,$$

$$a>R_{\min}, \tag{8}$$

where we have used the orthogonality relations

$$\int_{-\infty}^{+\infty}e^{i(k_z-k_z')z}dz=2\pi\delta(k_z-k_z') \tag{9}$$

and

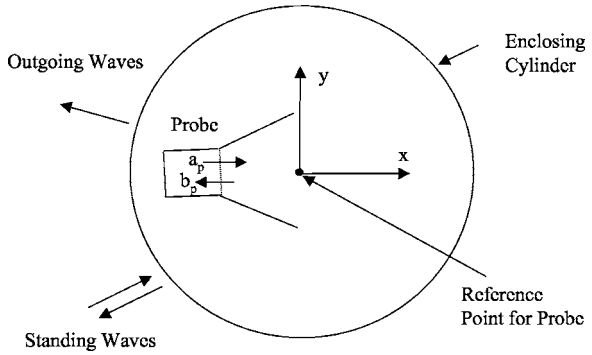


FIG. 2. Probe situated in a probe coordinate system. The input and output of the probe are determined by the mode amplitudes a_p and b_p, respectively. The total field outside the enclosing cylinder for the probe is described in terms of outgoing and standing waves.

$$\int_0^{2\pi} e^{i(n-n')\phi} d\phi = \begin{cases} 2\pi, & n = n' \\ 0, & n \neq n'. \end{cases} \tag{10}$$

Even though the radius a of the scan cylinder appears on the right-hand side of the expression (8) for spectrum, the expression (8) is independent of a. The formulas (5), (6), and (8) express the field everywhere outside a minimum cylinder that encloses the sources in terms of the field on the scan cylinder.

B. Far field in terms of the probe output on the scan cylinder

We shall next present a simple straightforward derivation of the probe-corrected formulas that express the field outside the scan cylinder in terms of the output of a known probe on the scan cylinder. We shall employ the outgoing and standing wave basis functions of the source scattering matrix formulation. The schematic for the cylindrical near-field measurement system is shown in Fig. 1. First we characterize the probe with respect to cylindrical waves.

1. Characterization of the probe

Start by defining a probe coordinate system in which the probe is fixed as shown in Fig. 2. The origin of this coordinate system is the reference point of the probe. The orientation of the probe with respect to a global coordinate system is defined in terms of the orientation of the probe coordinate system with respect to the global coordinate system. For cylindrical scanning we assume that the z axis of the probe coordinate system is parallel to the z axis of the global coordinate system that is used to describe the scanning geometry.

The probe is attached to a waveguide that supports only one propagating mode. The input to the probe is determined by the mode amplitude a_p and the output of the probe is determined by the mode amplitude b_p. We refer to b_p as the probe output and assume that the probe waveguide is perfectly matched ($a_p=0$) when the probe measures the field of the source.

The field outside the cylinder in Fig. 2 that encloses the probe can be described in terms of the outgoing waves

$$p^{(3)}(\rho,\phi,z) = H_n^{(1)}(k_\rho\rho)e^{in\phi}e^{ik_z z} \tag{11}$$

and the standing waves

$$p^{(1)}(\rho,\phi,z) = J_n(k_\rho\rho)e^{in\phi}e^{ik_z z}, \tag{12}$$

where $J_n(k_\rho\rho)$ is the Bessel function of order n. The probe receiving coefficients can now be defined as follows. If the total field outside the enclosing cylinder is the single standing cylindrical wave $J_n(k_\rho\rho)e^{in\phi}e^{ik_z z}$, the probe output is given by $b_p=C_n(kz)$, where $C_n(kz)$ are the probe receiving coefficients.

The probe receiving coefficients will now be expressed in terms of the probe's plane-wave receiving characteristic $R_p(k_x,k_y)$, defined to be the probe output when the incident field is the plane-wave $e^{i(k_x x+k_y y+k_z z)}$, with $k_x^2+k_y^2+k_z^2=k^2$. The plane-wave receiving characteristic, which is also defined with the probe located in the probe coordinate system, can be expressed both in terms of the plane-wave transmitting spectrum and the far-field pattern of the probe when the probe is a reciprocal electroacoustic transducer.[30,5] For example, if the probe is reciprocal and its far field is expressed in terms of a far-field pattern $\mathcal{F}_p(\theta,\phi)$ as $p(\mathbf{r})\sim\mathcal{F}_p(\theta,\phi)e^{ikr}/r$, the receiving characteristic is

$$R_p(k\cos\phi\sin\theta, k\sin\phi\sin\theta) = -\frac{\mathcal{F}_p(\pi-\theta,\pi+\phi)}{i\omega Y_p \rho_{m0}}, \tag{13}$$

where Y_p is the characteristic admittance for the propagating mode of the probe waveguide feed and p_{m0} is the mass density of the undisturbed fluid.

With the angles ϕ_k and θ_k defined such that $k_x=k_\rho\cos\phi_k$, $k_y=k_\rho\sin\phi_k$, and $k_z=k\cos\theta_k$, the cylindrical expansion of the incident plane-wave $e^{i(k_x x+k_y y+k_z z)}$ is

$$e^{i(k_x x+k_y y+k_z z)} = \sum_{n=-\infty}^{+\infty} i^n J_n(k_\rho\rho)e^{in(\phi-\phi_k)}e^{ik_z z}. \tag{14}$$

Note that $k_x<0$ and $\pi/2<\phi_k<3\pi/2$ for plane-wave components generated by sources that are located in the half space $x>0$. By inspection, the plane-wave receiving characteristic is found to be

$$R_p(k_x,k_y) = \sum_{n=-\infty}^{+\infty} i^n C_n(k_z)e^{-in\phi_k}, \tag{15}$$

which in turn yields the desired expression for the receiving coefficients

$$C_n(k_z) = \frac{i^{-n}}{2\pi}\int_0^{2\pi} R_p(k_\rho\cos\phi_k, k_\rho\sin\phi_k)e^{in\phi_k}d\phi_k. \tag{16}$$

As will be demonstrated below, the far field of the source can be determined from the plane-wave receiving characteristic in the region of propagating waves where k_x, k_y, and k_z are real.

2. Piston probe

We consider now a baffled circular receiving piston transducer whose piston resides in the y-z plane of the probe coordinate system. For brevity we shall refer to this receiv-

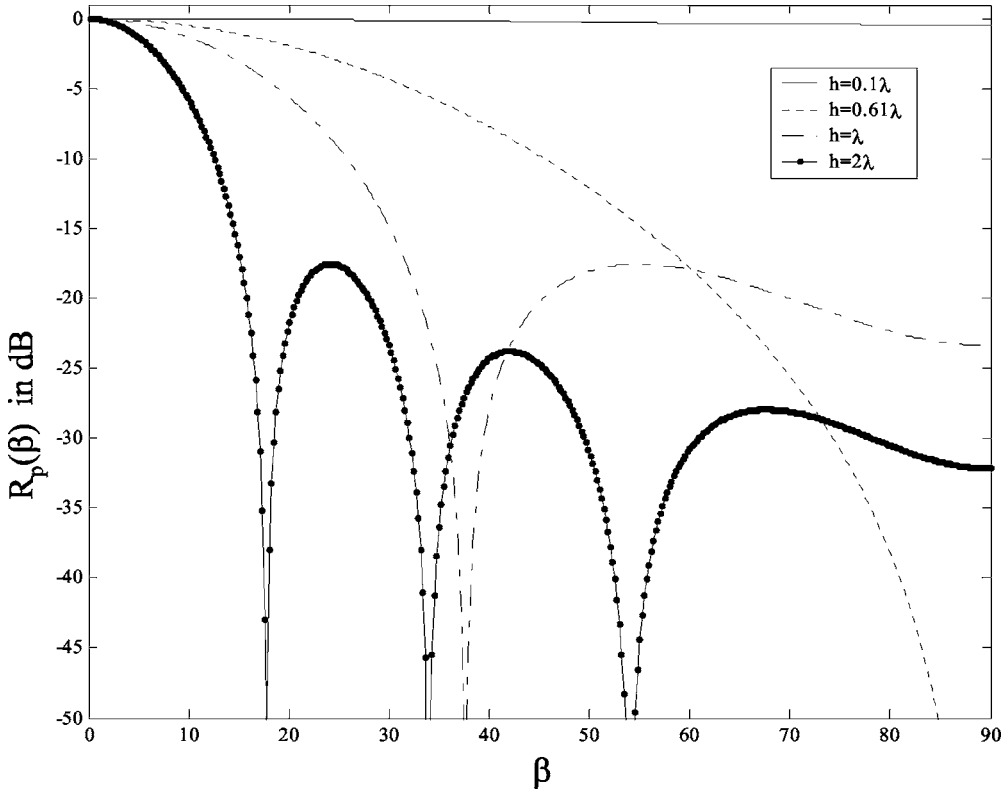


FIG. 3. Receiving pattern of piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$.

ing transducer as a piston probe. When the piston has radius h and is centered on the origin, the receiving characteristic takes the form

$$R_p(\beta)=\frac{2J_1(hk\sin\beta)}{hk\sin\beta}, \tag{17}$$

where β is the angle between the axis of the piston (x axis) and the direction of propagation of the incoming plane wave. We have normalized the receiving characteristic so that $R_p(0)=1$ for all h.

The receiving characteristic of the probe depends strongly on the radius of the piston as illustrated in Fig. 3, where $R_p(\beta)$ is plotted for $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$. For $h\ll\lambda$ the receiving characteristic is approximately one and the probe output simply equals the field (ideal probe). The first nonzero null of the Bessel function of order one is at 3.83, so for $hk=3.83$ the receiving characteristic is zero for $\beta=90°$. Hence, for $h<0.61\lambda$ the pattern has no nulls for real β (propagating modes) and for $h>0.61\lambda$ the pattern has at least one null for real β. We shall see that nulls of the receiving characteristic strongly affect the performance of the probe in near-field scanning applications.

By noting that $k\sin\beta=\sqrt{k_y^2+k_z^2}=\sqrt{k^2-k_x^2}$, the receiving characteristic as a function for k_x and k_y is found to be

$$R_p(k_x,k_y)=\frac{2J_1(h\sqrt{k^2-k_x^2})}{h\sqrt{k^2-k_x^2}} \tag{18}$$

and from Eq. (16) we get the following expression for the receiving coefficients:

$$C_n(k_z)=\frac{4}{\pi i^n}\int_0^{\pi/2}\frac{J_1(h\sqrt{k^2\sin^2\phi_k+k_z^2\cos^2\phi_k})}{h\sqrt{k^2\sin^2\phi_k+k_z^2\cos^2\phi_k}}\times\cos(n\phi_k)d\phi_k,\quad n\text{ even} \tag{19}$$

and $C_n(k_z)=0$ for n odd. Figure 4 shows the normalized receiving coefficients of the piston for four values of the piston radius and $k_z=0$. For $h\ll\lambda$, the receiving coefficients equal those of an ideal probe: $C_n=0$ for $n\neq 0$ and $C_0=1$. For large h, the modes with $|n|<kh$ are nonzero.

We shall also need to compute the probe output from a point source located in the probe coordinate system at $\mathbf{r}_1=x_1\hat{\mathbf{x}}+y_1\hat{\mathbf{y}}+z_1\hat{\mathbf{z}}$, with $x_1>0$. The probe output can be expressed conveniently as an integration over plane waves involving the probe receiving characteristic and the spectrum of the point source.[4,5] The plane-wave expansion of a point source is obtained from the Weyl identity[31,5]

$$\frac{e^{ik|\mathbf{r}-\mathbf{r}_1|}}{|\mathbf{r}-\mathbf{r}_1|}=\frac{i}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{e^{i[k_y(y-y_1)+k_z(z-z_1)+\sqrt{k^2-k_y^2-k_z^2}|x-x_1|]}}{\sqrt{k^2-k_y^2-k_z^2}}dk_ydk_z, \tag{20}$$

where $\sqrt{k^2-k_y^2-k_z^2}$ has non-negative real and imaginary parts. We insert the expressions for the point-source spectrum and the probe receiving characteristic into a plane-wave integral, write the integral in polar coordinates, and evaluate the angular integral in closed form in terms of Bessel functions, to get the following expression for the probe output

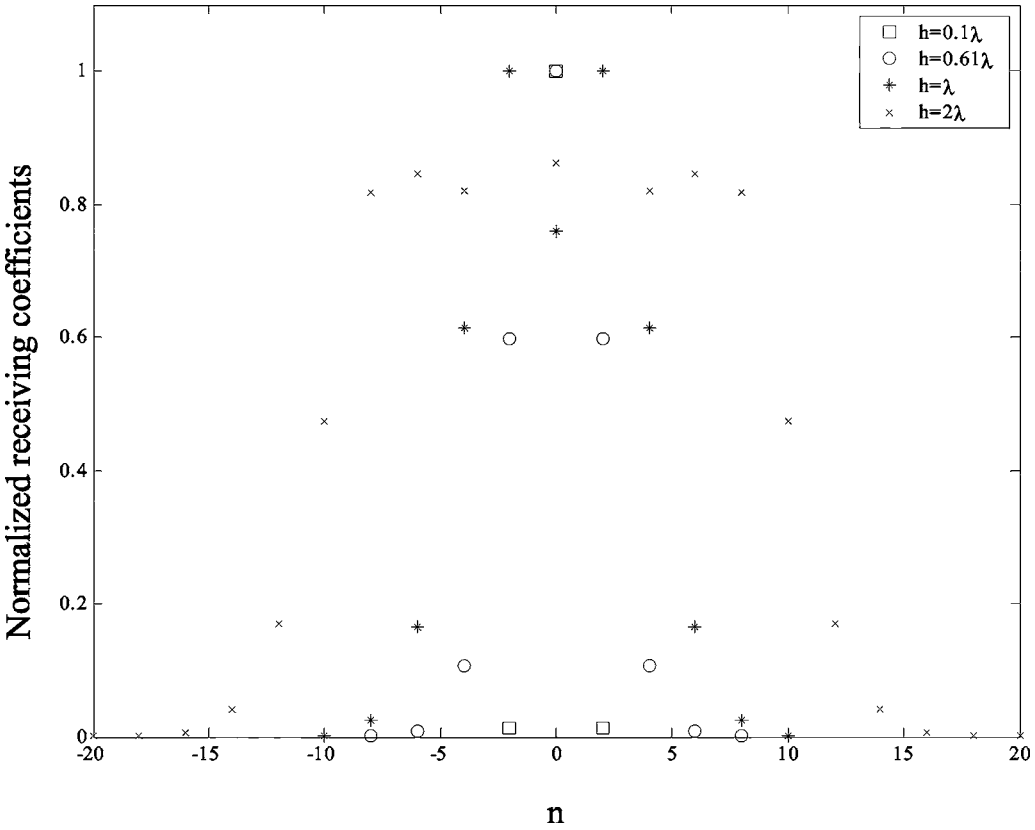


FIG. 4. Receiving coefficients at $k_z=0$ of piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$.

$$b_p = \frac{2i}{h}\int_0^{+\infty} u\frac{e^{ix_1\sqrt{k^2-u^2}}}{\sqrt{k^2-u^2}}\left[\frac{J_1(hu)}{u}J_0(u\sqrt{y_1^2+z_1^2})\right]du, \quad x_1>0, \tag{21}$$

where $\sqrt{k^2-u^2}$ has non-negative real and imaginary parts. The formula (21) is valid when the piston surface occupies the space $x=0$, $\sqrt{y^2+z^2}<h$ in the probe coordinate system in Fig. 2. The integrable singularity at $u=k$ in Eq. (21) may be removed by writing Eq. (21) as

$$b_p = \frac{2i}{h}\int_0^{+\infty} u\frac{e^{ix_1\sqrt{k^2-u^2}}}{\sqrt{k^2-u^2}}\left[\frac{J_1(hu)}{u}J_0(u\sqrt{y_1^2-z_1^2}) - \frac{J_1(hk)}{k}J_0(k\sqrt{y_1^2+z_1^2})\right]du + \frac{2e^{ikx_1}}{hx_1}\frac{J_1(hk)}{k}J_0(k\sqrt{y_1^2+z_1^2}), \quad x_1>0, \tag{22}$$

where the Weyl identity (20) has been used to calculate the last term. The formula (21) could also be obtained by invoking reciprocity to show that the probe output equals the field radiated by the piston at the location of the point source. The field of the circular piston was obtained by King,[32] and Eq. (21) is known as the King integral.[33]

Alternatively, the probe output can be expressed in terms of an integral of the incident pressure over the piston surface[34]

$$b_p = \frac{1}{\pi h^2}\iint_{y^2+z^2<h^2}\frac{e^{ik\sqrt{x_1^2+(y-y_1)^2+(z-z_1)^2}}}{\sqrt{x_1^2+(y-y_1)^2+(z-z_1)^2}}dydz. \tag{23}$$

Remarkably, the exact output of a baffled receiving piston transducer of arbitrary cross section is proportional to the integral over the piston surface of the incident pressure.[34] In yet another alternative representation, the output is expressed as a linear spatial differential operator acting on the incident pressure at the reference point of the probe.[16,18]

The formulas (21)–(23) hold for any piston size and any point source location with $x_1>0$, even when the point source is in the near field of the piston. When the point source is in the far field of the piston, its field can be approximated by a single plane wave in the region of the piston. From the expression (17) for the receiving characteristic, we obtain the far-field approximation for the probe output

$$b_p \sim \frac{2J_1(hk\sin\beta)}{hk\sin\beta}\frac{e^{ik|\mathbf{r}_1|}}{|\mathbf{r}_1|}, \quad \sin\beta = \frac{\sqrt{y_1^2+z_1^2}}{|\mathbf{r}_1|}. \tag{24}$$

Numerical calculations reveal that the error of the far-field formula (24) is 27%, 13%, and 3% when $h=\lambda$ and the point source is at $\mathbf{r}_1=5\lambda\hat{\mathbf{x}}+\lambda\hat{\mathbf{y}}$, $\mathbf{r}_1=5\lambda\hat{\mathbf{x}}+7\lambda\hat{\mathbf{y}}$, and $\mathbf{r}_1=30\lambda\hat{\mathbf{x}}+10\lambda\hat{\mathbf{y}}$, respectively. Hence, even with a relatively small piston, the exact formula is required to accurately compute the probe output from point sources less than 30λ away.

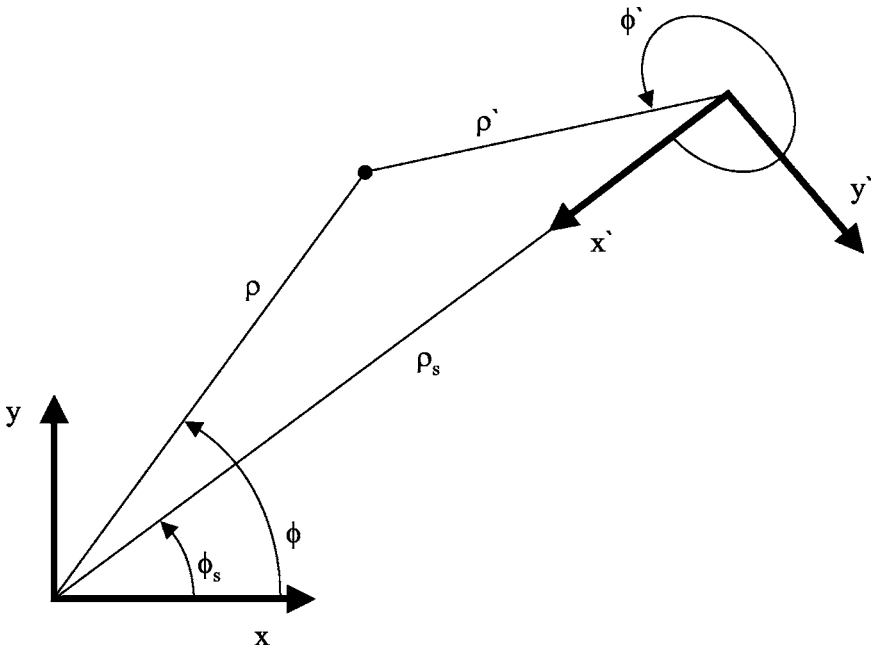


FIG. 5. Geometry for translation and rotation of cylindrical wave functions.

3. Probe-corrected formulas

The probe-corrected formulas for the field of the source will be derived in this section using a straightforward approach that involves two steps. First, an expression is derived for the probe output $b_p(a,\phi,z)$ in terms of the unknown expansion coefficients $F_m(k_z)$ for the source and the known probe receiving coefficients $C_n(k_z)$. Second, this expression for the probe output is inverted to get the expansion coefficients of the source in terms of the probe receiving coefficients and the probe output.

Multiple interactions between the source and the probe will be neglected. That is, the field scattered by the probe, rescattered by the source, and returned to the probe, does not change the output of the probe. The effect of multiple interactions can be taken into account formally by the source scattering matrix formulation.[8] However, this formulation does not provide quantitative information about these multiple interactions, and to obtain useful probe-corrected formulas they have to be neglected.

To derive an expression for the probe output when the probe is illuminated by the field of the source, we need to express the cylindrical waves in Eq. (5) in terms of cylindrical waves defined in the probe coordinate system. To distinguish between the global coordinate system (in which the scan cylinder is given by $\rho=a$) and the probe coordinate system (in which the probe is fixed) we attach primes to all coordinates defined with respect to the probe coordinate system.

Assume that the reference point of the probe is at (a,ϕ_s,z_s) on the scan cylinder. Then we need to transform the cylindrical wave functions in Eq. (5) into cylindrical wave functions in the probe coordinate system, which is centered at (a,ϕ_s,z_s) with its x axis pointing toward the center of the global coordinate system. To achieve this transformation we use the rotation-translation formula for cylindrical wave functions

$$H_m^{(1)}(\rho)e^{im\phi} = (-1)^m e^{im\phi_s}\sum_{n=-\infty}^{+\infty} H_{n-m}^{(1)}(\rho_s)J_n(\rho')e^{in\phi'}, \quad \rho_s > \rho', \tag{25}$$

where the primed quantities are defined in the probe coordinate system and the unprimed quantities are defined in the global coordinate system, as shown in Fig. 5.

Equation (25) can be inserted into expression (5) to get the following formula for the pressure on the scan cylinder:

$$p(a,\phi_s,z_s) = \sum_{m=-\infty}^{+\infty}(-1)^m e^{im\phi_s} \times \sum_{n=-\infty}^{+\infty}\int_{-\infty}^{+\infty}[J_n(k_\rho\rho')e^{in\phi'}e^{ik_z z'}]F_m(k_z)H_{n-m}^{(1)} \times (k_\rho a)e^{ik_z z_s}dk_z, \tag{26}$$

which expresses the outgoing waves that emanate from the source in the global coordinate system in terms of standing waves $J_n(k_\rho\rho')e^{in\phi'}e^{ink_z z'}$ in the probe coordinate system. From the definition of the probe receiving coefficients, we see that the total probe output is

$$b_p(a,\phi,z) = \sum_{m=-\infty}^{+\infty}(-1)^m e^{im\phi_s}\sum_{n=-\infty}^{+\infty}\int_{-\infty}^{+\infty} C_n(k_z)F_m(k_z)H_{n-m}^{(1)} \times (k_\rho a)e^{ik_z z_s}dk_z. \tag{27}$$

The expression (27) is easily inverted by use of the orthogonality relations (9) and (10) to get the final probe-corrected expression for the spectrum

$$F_m(k_z) = D_m(k_z)I_m(k_z), \tag{28}$$

where

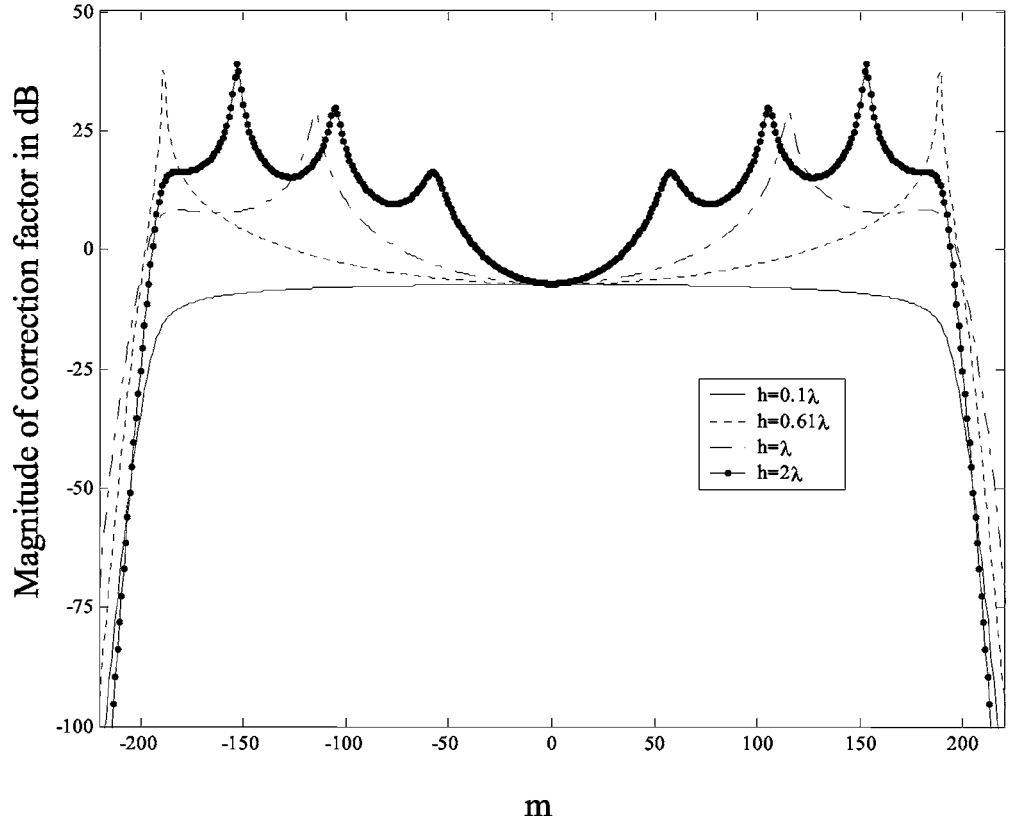


FIG. 6. Correction factor at $k_z=0$ for piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$ that measure the field on a scan cylinder of radius $a-30\lambda$.

$$I_m(k_z)=\int_{-\infty}^{+\infty}\int_0^{2\pi} b_p(a,\phi_s,z_s)e^{-im\phi_s}e^{-ik_z z_s}d\phi_s dz_s \tag{29}$$

is a Fourier integral of the probe output over the scan cylinder and

$$D_m(k_z)=\frac{(-1)^m}{4\pi^2\sum_{n=-\infty}^{+\infty} C_n(k_z)H_{n-m}^{(1)}(k_\rho a)} \tag{30}$$

is a correction factor that depends on the probe receiving coefficients and the scan cylinder radius. The far field of the source is expressed in Eq. (6) in terms of the expansion coefficients $F_m(k_z)$ evaluated at $k_z=k\cos\theta$, and the formula (16) determines the probe receiving coefficients $C_n(k_z)$ in terms of the plane-wave receiving characteristic of the probe. For an ideal probe whose output equals the pressure ($C_n=0$ for $n\neq 0$ and $C_0=1$), Eq. (28) reduces correctly to the nonprobe-corrected formula (8).

Figure 6 shows the magnitude in decebels of the correction factor (30) for piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$. The scan cylinder radius is $a=30\lambda$ and the observation direction is $\theta=90°$, corresponding to $k_z=0$. We shall now describe how the correction factor corrects for the probe pattern for far-field observation points near $\theta=\pi/2$.

First assume that the field is produced by a point source at the origin so that $R_{\min}=0$. Then the expansion coefficients $F_m(0)$ are zero for $m\neq 0$, and only the correction factor $D_0(0)$ comes into play. For points on the scan cylinder near the x-y plane, the output is roughly the same for all four probes because the point source is in the direction of the main beam (see Fig. 3 with $\beta=0$). The far field of the point source at $\theta=\pi/2$ is mainly determined by the probe output near the x-y plane, so the correction factor $D_0(0)$ should be roughly the same for the four probes, as confirmed by Fig. 6.

Next assume that the field is produced by a point source on the x axis at $x=3a/4$, for which the expansion coefficients $F_m(0)$ with $|m|<150$ are non-negligible. The probe sees this point source at an angle $\beta=37°$ when the probe is at $(x,y,z)=(0,a,0)$ and $(x,y,z)=(0,-a,0)$. At all other observation points in the x-y plane, the probes sees the point scatterer at an angle β that is less than 37°. Consider the piston probe with $h=\lambda$. From Fig. (3) we see that $\beta=37°$ is near the first null of this probe, so very low outputs are obtained when the probe is near $(x,y,z)=(0,a,0)$ and $(x,y,z)=(0,-a,0)$. To compensate for these low outputs, the correction factor $D_m(0)$ has a peak at $|m|=120$ and attains large values around this peak. Hence, to compute some of the higher-order expansion coefficients of the source $F_m(0)$, the integral over the scan cylinder $I_m(0)$ must be boosted by a large correction factor $D_m(0)$. These arguments can be repeated for the other piston probes in Fig. 3.

This discussion shows that nulls and low sidelobe levels in the probe receiving pattern result in large values of the correction factor. Correction factors that vary significantly with m, can often lead to errors in the computed far field because some of the higher-order expansion coefficients of the source F_m are computed as the product of a small number that can be relatively inaccurate (the integral of the probe output over the scan cylinder I_m) and a very large number

(the correction factor D_m). Therefore, it is advantageous to avoid probes that have nulls and low sidelobe levels in the region of the sources.

In Sec. IV we show how the far field can be computed from sampled values of the probe output on the scan cylinder. If the angular sampling rate is $\Delta\phi$, the far-field summation (6) includes the terms with $|n| < \pi/\Delta\phi$ (recall that the spectrum in Eq. (6) is given in Eq. (28) as $F_n = D_n I_n$). Hence, by decreasing the sample spacing $\Delta\phi$ the far-field error may increase because more of the peaks of the correction factor are included. This phenomenon is illustrated in Sec. V. Of course, the sample rate $\Delta\phi$ must always be chosen small enough to properly resolve the field of the source.

We shall now briefly discuss an approximate probe-corrected formula obtained by inserting[35]

$$H_{n-m}^{(1)}(k_\rho a) \simeq (-1)^m (-i)^n e^{-in \arcsin(m/[k_\rho a])} H_m^{(1)}(k_\rho a) \quad (31)$$

into Eq. (30) to get

$$D_m(k_z) \simeq \frac{1}{4\pi^2 R_p(k_p \cos[\arcsin(m/[k_\rho a]) + \pi], k_\rho \sin[\arcsin(m/[k_\rho a]) + \pi]) H_m^{(1)}(k_\rho a)}. \quad (32)$$

An electromagnetic analog of Eq. (32) was derived by Borgiotti[9,10] and subsequently rederived by Yaghjian.[8] The approximate formula (32) expresses the correction factor $D_m(k_z)$ directly in terms of the receiving characteristic $R_p(k_x, k_y)$ and avoids the use of the cylindrical receiving coefficients $C_n(k_z)$ that occur in the exact formula (30). By invoking the relation (13), the formula (32) expresses the correction factor directly in terms of the probe far-field pattern. For $k_\rho a \gg 1$, expression (31) is accurate only for small $|n|$ and values of $|m|$ that are not too close to $k_\rho a$. Hence, Eq. (32) is accurate only for small probes with only a few nonzero receiving coefficients $C_n(k_z)$ and for sources with $R_{\min}$ not too close to a. The approximate formula (32) is singular when $\arcsin(m/[k_\rho a]) + \pi$ coincides with a null of the receiving characteristic. These singularities correspond to the "peaks" discussed above where the exact correction factor attains large values but remains finite. The approximate formula (32) thus fails to correctly predict the values of the correction factor near its peaks. However, it is a valuable formula that avoids the use of the cylindrical receiving coefficients and produces accurate far fields for small probes whose patterns are null free. As we shall demonstrate below, small probes have additional advantages.

The fact that the exact correction factor is without singularities distinguishes cylindrical scanning from planar scanning: In planar scanning the exact correction factor is proportional to $1/R_p$ and thus has singularities at the nulls of R_p, as proven by Hansen and Yaghjian[5]. This discussion was confined to far-field observation directions near $\theta = \pi/2$. In Sec. V we shall show how nulls in the probe pattern affects the computed far field for observation points away from $\theta = \pi/2$.

III. TRUNCATION OF THE SCAN CYLINDER

The exact expressions (8) and (28)–(30) for the expansion coefficients $F_m(k_z)$ involve an integration over an entire infinite scan cylinder of either the pressure or the probe output. Practical near-field measurements are carried out only over a truncated scan cylinder, which we assume is given by $-z_0 \leqslant z \leqslant z_0$ and $\phi_0 \leqslant \phi \leqslant 2\pi - \phi_0$, where $z_0 > 0$ and $0 \leqslant \phi_0 < \pi$. The near-field data required by Eqs. (8) and (28)–(30) is therefore not available in practice. Fortunately, experience has shown that very accurate far fields can be obtained in certain regions from measurements on a truncated scan cylinder. In this section we shall study the errors caused by truncation and develop methods for reducing them. Truncation in the z and ϕ directions are considered in Secs. III A and III B, respectively.

First we state a result from Hansen *et al.*[22] that restricts the range of observation angles for which we can accurately compute the far field from a truncated scan: *If the scanned area is thought of as a transparent surface and the rest of the infinite scan cylinder is opaque, then accurate far fields can be achieved only in directions from which one can see the entire source region*. Similar results have been cited for planar and spherical scanning surfaces.[36–39]

A. Truncation in the *z* direction

Typically the field outside the scanned area is neglected and the exact expression (29) is approximated by

$$I_n(k_z) \simeq \int_{-z_0}^{z_0} \int_{\phi_0}^{2\pi-\phi_0} b_p(a, \phi_s, z_s) e^{-in\phi_s} e^{-ik_z z_s} d\phi_s dz_s. \quad (33)$$

It is thus assumed that the contribution to the integral (29) from the region not covered by the scan is negligible. For cylindrical near-field measurements of directive transducers this assumption is often valid because the measurements are set up so that very little energy is radiated in directions not covered by the truncated scan cylinder. However, a nondirective transducer or a scatterer may radiate significantly in all directions, including the ones that are not covered by the scan cylinder. Hence, it is not always a valid approximation to neglect the contribution to the integral in Eq. (29) from the region not covered by the scan.

We shall next present two sets of asymptotic correction formulas that approximate the contribution from the region of the scan cylinder with $|z| > z_0$. Detailed derivations can be

found in Hansen *et al.*[22] First we assume that the field on the scan cylinder behaves roughly as a plane wave $e^{ik|z|}$ as $z \to \pm\infty$ and find that

$$I_n(k_z) \simeq \int_{-z_0}^{z_0}\int_{\phi_0}^{2\pi-\phi_0} b_p(a,\phi_s,z_s)e^{-in\phi_s}e^{-ikz_z s}d\phi_s dz_s - \frac{e^{ikz_0}}{i(k+k_z)}\int_{-\phi_0}^{2\pi-\phi_0} b_p(a,\phi_s,-z_0)e^{-in\phi_s}d\phi_s - \frac{e^{-ikz_0}}{i(k-k_z)}\int_{\phi_0}^{2\pi-\phi_0} b_p(a,\phi_s,z_0)e^{-in\phi_s}d\phi_s. \quad (34)$$

We use the term "edge-corrected" formula because the integral over the uncovered areas of the infinite scan cylinder are approximated by integrations along the top and bottom edges of the truncated scan cylinder.

The second set of correction formulas assumes that field behaves as a spherical wave e^{ikr} as $z \to \pm\infty$. The edge-corrected formula for the spherical-wave assumption is[40]

$$I_n(k_z) \simeq \int_{-z_0}^{z_0}\int_{\phi_0}^{2\pi-\phi_0} b_p(a,\phi_s,z_s)e^{-in\phi_s}e^{-ikz_z s}d\phi_s dz_s - \frac{e^{ikz_0}}{i(k\cos\theta_0+k_z)}\int_{\phi_0}^{2\pi-\phi_0} b_p(a,\phi_s,-z_0)e^{-in\phi_s}d\phi_s - \frac{e^{-ikz_0}}{i(k\cos\theta_0-k_z)}\int_{\phi_0}^{2\pi-\phi_0} b_p(a,\phi_s,z_0)e^{-in\phi_s}d\phi_s, \quad (35)$$

where $\cos\theta_0 = z_0/\sqrt{a^2+z_0^2}$. Note that the edge-corrected formula (35) is singular for far-field observation directions with $\theta=\theta_0$ and $\theta=\pi/2-\theta_0$. This follows from the far-field expression (6) which shows that the far field in the direction (θ,ϕ) is determined from the spectrum evaluated at $k_z = k\cos\theta$. The edge-corrected formula (34) does not have these singularities.

B. Truncation in the ϕ direction

For the angular truncation of the scan cylinder we cannot derive asymptotic correction formulas similar to Eqs. (34) and (35) because no asymptotic expression suggests itself for the ϕ dependence of the field. Indeed, the ϕ dependence of the field on the scan cylinder depends strongly on the source shape, so we do not have a general expression for $b_p(a,\phi,z)$ that holds when ϕ is outside the scanned region $\phi_0 \leqslant \phi \leqslant 2\pi-\phi_0$.

To determine the effect of ϕ truncation, insert the formula (28) for the spectrum $F_m(k_z)$ into the far-field formula (6) to get

$$p(r,\theta,\phi) \sim \frac{e^{ikr}}{r}\int_{-\infty}^{+\infty}\int_0^{2\pi} b_p(a,\phi_s,z_s)G(k,a,\theta,\phi-\phi_s;\{C_n\})e^{-ikz_s\cos\theta}d\phi_s dz_s, \quad (36)$$

where

$$G(k,a,\theta,\phi;\{C_n\}) = \sum_{m=-\infty}^{+\infty}\frac{(-1)^m e^{-i(m+1)\pi/2}e^{im\phi}}{2\pi^2\sum_{n=-\infty}^{+\infty} C_n(k\cos\theta)H_{n-m}^{(1)}(ka\sin\theta)} \quad (37)$$

will be called the taper function. Equation (36) expresses the far field of the source directly in terms of an integration over the scan cylinder of the probe output multiplied by the taper function. This expression is useful for determining the ϕ-truncation errors but not efficient for actually computing the far field since the fast Fourier transform (FFT) cannot be fully exploited. We shall now examine the taper function (37) for an ideal probe and for piston probes.

1. Ideal probe

For an ideal probe ($C_n=0$ for $n\neq 0$ and $C_0=1$), the taper function (37) reduces to

$$G_i(ka,\theta,\phi) = \sum_{m=-\infty}^{+\infty}\frac{e^{-i(m+1)\pi/2}e^{im\phi}}{2\pi^2 H_m^{(1)}(ka\sin\theta)}. \quad (38)$$

The piston probe with $h=0.1\lambda$ is approximately an ideal probe. By inspection, we see that the function $G_i(ka,\theta,\phi-\phi_s)e^{-ikz_s\cos\theta}$ equals the total field on a hard circular cylinder that is illuminated by a plane wave with amplitude $(2\pi i)^{-1}ka\sin\theta$ that propagates in the $-\hat{\mathbf{r}}$ direction. Here, $\hat{\mathbf{r}}$ is the unit vector for the far-field observation direction. Hence, *the far field of the source is obtained by integrating over the scan cylinder the product of the near field of the source and the total field on a hard cylinder, which coincides with the scan cylinder and is illuminated by a plane wave that originates at the far-field observation point*.

Jones[41] used the Watson transform to prove that as $ka\sin\theta\to\infty$ the function $G_i(ka,\theta,\phi)$ decays exponentially for $\pi/2<|\phi|\leqslant\pi$, has a transition zone around $|\phi|=\pi/2$, and behaves as

$$G_i(ka,\theta,\phi) \sim \left[\frac{ka\sin\theta}{2\pi i}\right]\cos\phi e^{-ika\sin\theta\cos\phi}, \quad |\phi|<\pi/2, \quad ka\sin\theta\to+\infty. \quad (39)$$

Figure 7 shows $G_i(ka,\theta,\phi)$ and its asymptotic approximation (39) for $ka\sin\theta=25$. The asymptotic approximation works well even at this low value of $ka\sin\theta$.

The qualitative behavior of $G_i(ka,\theta,\phi)$ plotted in Fig. 7 for $ka\sin\theta=25$ remains the same for all larger $ka\sin\theta$. Hence, the far-field formula (36) shows that the field in a given direction with angle ϕ is determined by integrating the field on the scan cylinder multiplied by an oscillating function of ϕ_s that is symmetric and decays away from $\phi_s=\phi$. In other words, the cylindrical wave-function expansion of the far field automatically applies an angular taper function to the field on the scan cylinder. No taper function is applied for the z_s dependence. The asymptotic expression (39) shows that the oscillations increase in frequency with increasing $ka\sin\theta$.

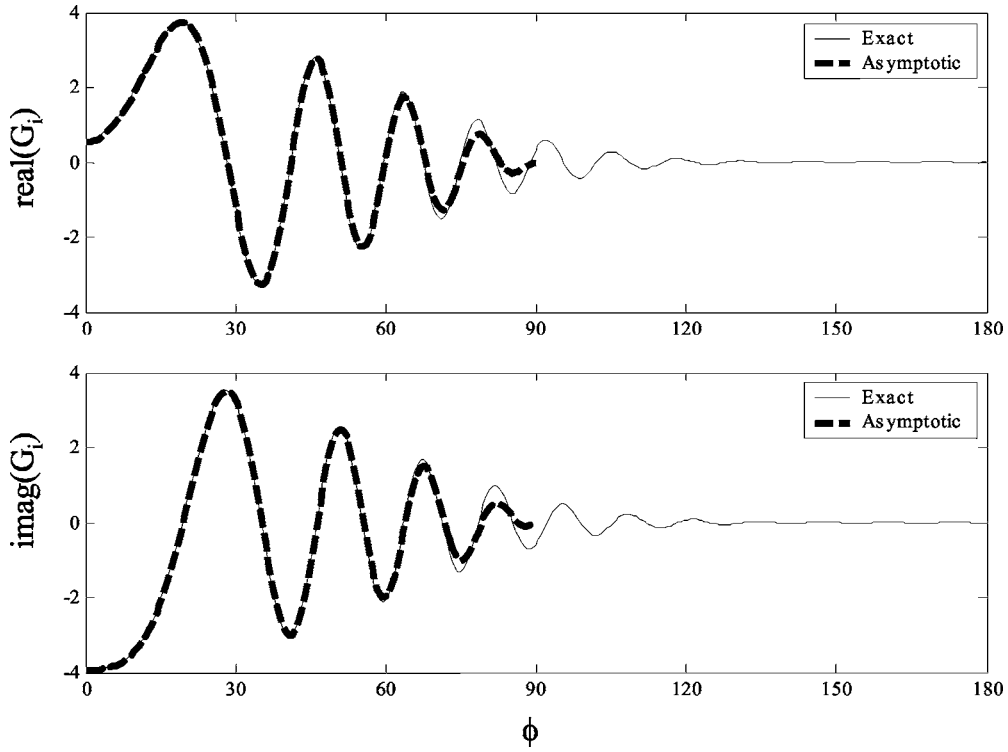


FIG. 7. Real (top) and imaginary (bottom) parts of the function G_i and its asymptotic approximation for $ka \sin \theta = 25$.

2. Piston probe

Figure 8 shows an upper envelope of the taper function in Eq. (37) for piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$. The scan cylinder radius is $a=30\lambda$ and the observation direction is $\theta=90°$, corresponding to $k_z=0$. To visualize the rapidly oscillating taper function, we computed an upper envelope that equals the maximum magnitude of the taper function over 5° intervals.

Let us first discuss the behavior of the taper function for $h=0.1\lambda$ (this piston probe is almost an ideal probe whose

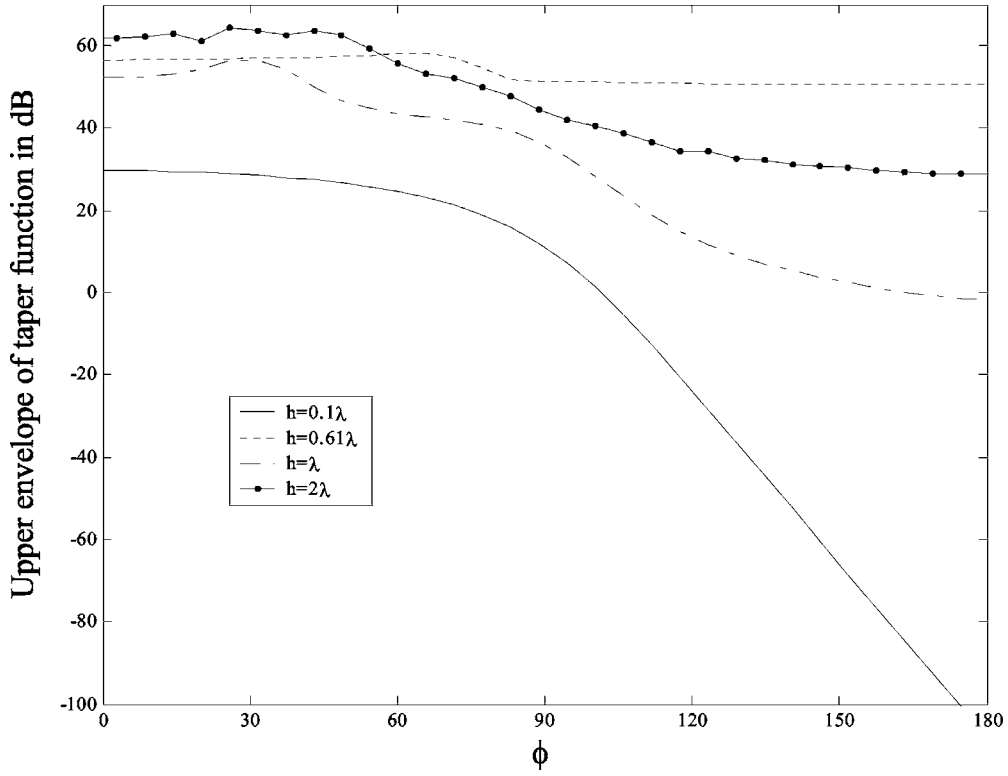


FIG. 8. Upper envelope of the taper functions for $a=30\lambda$ and $\theta=90°$ for piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$. All non-negligible terms of the expansion for the taper function are included.

output equals the pressure). In the region from $\phi=0°$ to $\phi=90°$ the taper function decays at a moderate rate from 30 to about 5 dB. Beyond $\phi=90°$ the decay is exponential (creeping-wave region), and at $\phi=180°$ the magnitude of the taper function is below −100 dB. Hence, the probe output on the scan cylinder at points that lie more than 90° away from the far-field observation direction are attenuated by more than 25 dB when the fields are measured with a piston probe of radius $h=0.1\lambda$. We shall later see that the strong taper helps reduce truncation effects.

The taper functions for the piston probes with $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$ do not exhibit such rapid decay. For example, the taper function for the piston probe of radius $h=0.61\lambda$ decays only 5 dB over the entire range from $\phi=0°$ to $\phi=180°$, and thus only a mild taper is applied to the probe output. The values of the taper functions in Fig. 8 were computed from the summation (37) with all non-negligible terms included. For practical far-field calculations, the summation in Eq. (37) is truncated at a mode number determined by the size of the source region. This truncation of the summation strongly affects the taper function, as will be demonstrated in Sec. V.

C. Green's function representation

Let us briefly discuss a far-field formula that is based on the free-space Green's function rather than the cylindrical wave functions. From Hansen *et al.*[5] it follows that the far field can be expressed in terms of the pressure and its normal derivative on the scan cylinder as

$$p(r,\theta,\phi) \sim -\frac{iae^{ikr}}{4\pi r}\int_{-\infty}^{+\infty}\int_{0}^{2\pi}\left[ik\hat{\mathbf{r}}\cdot\hat{\boldsymbol{\rho}}' p(\mathbf{r}') + \frac{\partial}{\partial\rho'}p(\mathbf{r}')\right]e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}d\phi' dz'. \tag{40}$$

The normal derivative of the pressure can be related to the velocity by the formula $\partial/\partial\rho'\, p(\mathbf{r}')=i\omega\rho_{m0}\boldsymbol{\rho}'\cdot\mathbf{v}(\mathbf{r}')$, where ρ_{m0} is the mass density of the undisturbed fluid.

This formula is seldom used in near-field scanning because it requires that both the pressure and its normal derivative be measured. Unlike the formulas that are based on the cylindrical wave functions, the formula (40) does not apply an angular taper function to the field on the scan cylinder. Hence, angular truncations of the scan cylinder would in general affect Eq. (40) more than they affect formulas based on cylindrical wave functions.

IV. SAMPLING THEOREMS AND COMPUTATION SCHEMES

The formulas derived so far for the far field of the source involve integrations of the probe output over the scan cylinder and infinite summations over angular modes. In this section we discretize these formulas so that the far field of the source can be computed from a finite number of probe-output values sampled at regular intervals on the scan cylinder. As part of the discretization, we present sampling theorems that determine the sample spacing required for accurate far-field calculation.

The scan cylinder is truncated and given in cylindrical coordinates by $\rho=a$ and $|z|\leq z_0$, and the field on the scan cylinder is measured at the grid points expressed in terms of the cylindrical coordinates as

$$\phi=(\ell-1)\Delta\phi, \quad \ell=1,2,\ldots,N_\phi, \tag{41}$$

and

$$z=(m-1)\Delta z - z_0, \quad m=1,2,\ldots,N_z. \tag{42}$$

The formulas of this section require the probe over the entire 360° range. If the scan cylinder is angularly truncated, one simply inserts zeros into the array positions that correspond to measurement points that are outside the scanned area. Note that z_0 satisfies the relation $z_0=(N_z-1)\Delta z/2$.

If every point on the scan cylinder is at least one wavelength or so away from the nearest point on a nonresonant source,[42] the sampling intervals $\Delta\phi$ and Δz can be chosen as follows:[17]

$$\Delta\phi=\frac{2\pi}{N_\phi}, \quad N_\phi=2\,\mathrm{int}(kR_{\min}+n_1), \tag{43}$$

$$\Delta z=\frac{\lambda}{2}, \tag{44}$$

where $R_{\min}$ is the radius of the minimum cylinder, n_1 is a small integer, and λ is the wavelength. In Eq. (43) the function "int" denotes the integer value.

The sampling theorems show that the infinite summation (6) can be replaced by the finite summation

$$p(r,\theta,\phi) \sim \frac{2e^{ikr}}{r}\sum_{n=-N_0}^{N_0} D_n(k\cos\theta)I_n(k\cos\theta)e^{in\phi}e^{-i\pi(n+1)/2}, \tag{45}$$

where $D_n(k\cos\theta)$ is the correction factor defined in Eq. (30), and N_0 is an integer that must be large enough to include all the significient modes of a give source region. When strong edge effects are present, N_0 should be large enough to prevent aliasing caused by the discontinuity of the near-field data at the truncation edges.[22] Since only the values of $I_n(k\cos\theta)$ with $|n|\leq N_\phi/2$ can be computed accurately, one must chose $N_0\leq N_\phi/2$.

Next we introduce the two-dimensional array

$$\tilde{b}_p(\ell,m)=b_p(a,(\ell-1)\Delta\phi,(m-1)\Delta z-z_0),$$

$$\ell=1,2,\ldots,N_\phi, \quad m=1,2,\ldots,N_z, \tag{46}$$

which contains the probe output b_p at the grid points. A straightforward discretization of the formula (33) for the spectrum then gives

$$I_n(k_z)=\Delta z\Delta\phi e^{ik_z z_0}\sum_{\ell=1}^{N_\phi}\sum_{m=1}^{N_z} e^{-in\Delta\phi(\ell-1)}e^{-ik_z\Delta z(m-1)}\tilde{b}_p(\ell,m). \tag{47}$$

Note that both the far-field formula (45) and the probe-corrected formula (47) for the spectrum can be computed efficiently with the FFT.

We can also discretize the edge-corrected formulas (34) and (35) to get

$$I_n(k_z) = \Delta z \Delta\phi e^{ik_z z_0} \sum_{\ell=1}^{N_\phi} \sum_{m=1}^{N_z} e^{-in\Delta\phi(\ell-1)} e^{-ik_z \Delta z(m-1)} \tilde{b}_p(\ell,m) - \frac{\Delta z \Delta\phi e^{ik_z z_0}}{2} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,1) - \frac{\Delta z \Delta\phi e^{-ik_z z_0}}{2} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,N_z) - \frac{\Delta\phi e^{ik_z z_0}}{i(k+k_z)} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,1) - \frac{\Delta\phi e^{-ik_z z_0}}{i(k-k_z)} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,N_z) \quad (48)$$

when the field is assumed to behave as $e^{ik|z|}$, and

$$I_n(k_z) = \Delta z \Delta\phi e^{ik_z z_0} \sum_{\ell=1}^{N_\phi} \sum_{m=1}^{N_z} e^{-in\Delta\phi(\ell-1)} e^{-ik_z \Delta z(m-1)} \tilde{b}_p(\ell,m) - \frac{\Delta z \Delta\phi e^{ik_z z_0}}{2} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,1) - \frac{\Delta z \Delta\phi e^{-ik_z z_0}}{2} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,N_z) - \frac{\Delta\phi e^{ik_z z_0}}{i(k\cos\theta_0+k_z)} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,1) - \frac{\Delta\phi e^{-ik_z z_0}}{i(k\cos\theta_0-k_z)} \sum_{\ell=1}^{N_\phi} e^{-in\Delta\phi(\ell-1)} \tilde{b}_p(\ell,N_z) \quad (49)$$

when the field is assumed to behave as e^{ikr} and $\cos\theta_0 = z_0/\sqrt{a^2+z_0^2}$. The second and third in Eqs. (48) and (49) ensure that the contributions from the top and bottom edges are included only once. The number of operations required to compute the spectrum from any one of the formulas (47)–(49) is on the order of $N_\phi N_z \log_2(N_\phi, N_z)$. Hence, the edge-correction terms do not significantly add to the computational effort.

V. FAR-FIELD ERRORS FOR A POINT SOURCE

In this section we shall compute the far field of an acoustic point source from near-field measurements on a truncated scan cylinder. The field is measured with piston probes of varying size and the probe output is computed from the exact formula (22). Hence, if the scan cylinder is untruncated, the computed far field equals the exact far field, regardless of the size of the piston probe. The acoustic field is generated by a single point source (i.e., a nondirective source) located inside the scan cylinder, so the truncation edges of the scan cylinder are illuminated by relatively strong fields, and the truncation effects will be clearly visible. The scan cylinder is given by $a=30\lambda$ and $-40\lambda < z < 39.5\lambda$. Throughout, the sampling spacing in z is $\Delta z = \lambda/2$. First assume that the point source is located at $\mathbf{r}_s = 12\lambda\hat{\mathbf{y}} - 5\lambda\hat{\mathbf{z}}$.

TABLE I.

Average far-field error over the region $0 \leq \phi < 2\pi$ for $\theta = \pi/2$				
	$h=0.1\lambda$ (%)	$h=0.61\lambda$ (%)	$h=\lambda$ (%)	$h=2\lambda$ (%)
No edge correction	3.1	0.7	0.5	0.3
Plane-wage edge correction	0.9	0.2	0.1	0.1
Spherical-wave edge correction	1.3	0.2	0.2	0.1

We begin by investigating the improvements in far-field accuracy achieved with the formulas that correct for the truncations at the top and bottom edges of the scan cylinder. We let the scan cylinder be untruncated in ϕ and chose $N_\phi = 360$ corresponding to an angular sample spacing of $\Delta\phi = 1°$. (With $R_{\min} = 12\lambda$, the required sample spacing is about 2.3°.) The sample spacing in z is $\Delta z = \lambda/2$. We shall compare the far fields obtained with four different probe sizes: $h = 0.1\lambda$, $h = 0.61\lambda$, $h = \lambda$, and $h = 2\lambda$. The exact far field of a single point source located at $\mathbf{r}_s$ is

$$p_0(\mathbf{r}) \sim e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}_s} \frac{e^{ikr}}{r}. \quad (50)$$

Table I shows the average error of the far field for $\theta = \pi/2$ and $0 \leq \phi < 2\pi$ computed from this z-truncated scan using Eqs. (47)–(49).

We see that the edge-correction formulas significantly reduce the far-field error for all sizes of the piston probe. Moreover, the error is smaller for directive probes because of the reduced magnitude of the probe output near the top and bottom edges of the scan cylinder. In the remainder of this section we employ the edge-corrected formula (48) based on the plane-wave assumption.

Figure 9 shows the far-field error as a function of θ for $\phi = 0$ with the scan cylinder still untruncated in ϕ. The error is calculated at the angles $\theta = \arccos(k_z/k)$ with increments of k_z/k equal to 0.0125 (these increments automatically come out of the FFT when the sample spacing in z is $\lambda/2$ and no zero padding is applied). The vertical solid lines in Fig. 9 are the limiting angles at which the line that begins at the point source and ends at the far-field observation point touches the upper and lower truncation edges of the scan cylinder. As expected, outside these limiting angles the far-field error is so large that the computed far field is useless.

The far-field error curves have spikes that correspond to nulls in the probe patterns (see Fig. 3). These spikes are not exactly symmetric because the point source is not in the x-y plane. They result from the fact that the probe output is very small at values of z for which the point source is near the direction of a null. When the far field is computed in the corresponding θ direction, the correction factor is large and thus the errors caused by z truncation lead to an inaccurate value for the far field.

We now compute the far field of the point source at $\theta = \pi/2$ from measurements over the $\phi - z$ truncated scanning

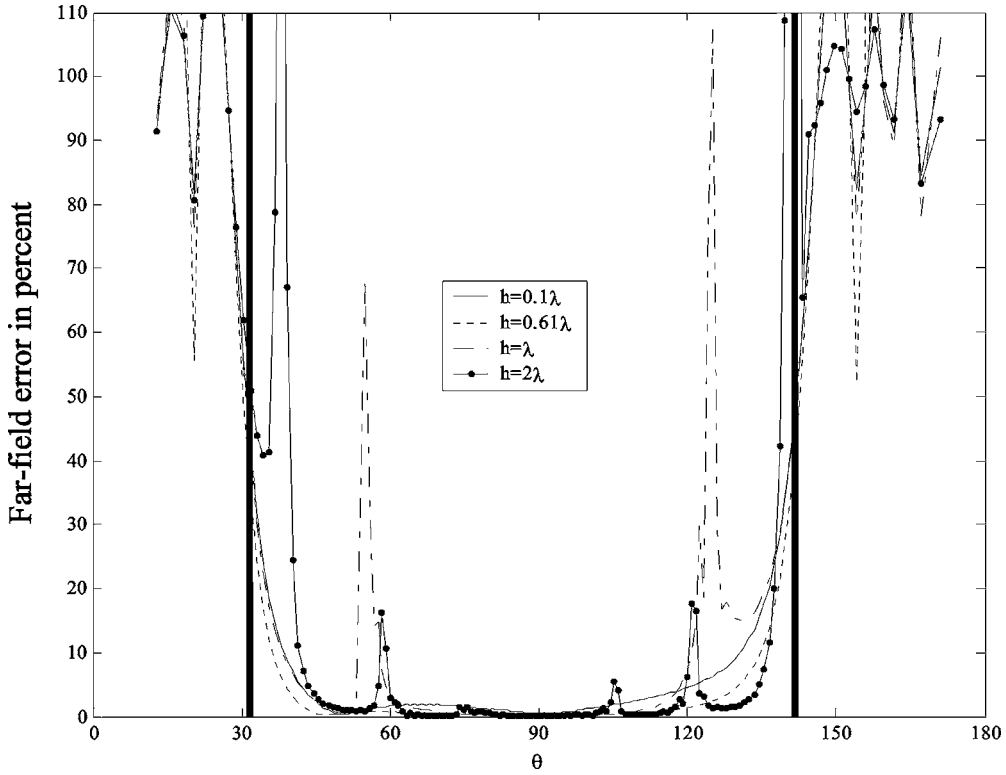


FIG. 9. Far-field error at $\phi=0°$ computed from a z-truncated scan with $a=30\lambda$, $0<\phi<360°$, and $-40\lambda<z<39.5\lambda$. The far-field error is calculated at $\theta=\arccos(k_z/k)$ with increments of k_z/k equal to 0.0125. The limiting angles in θ are indicated with vertical solid lines. The field is produced by a single point source at $(x,y,z)=(0,12\lambda,-5\lambda)$.

area given by $a=30\lambda$, $-40\lambda<z<39.5\lambda$, and $60°<\phi<300°$. Figure 10 shows this scanning area as a solid circular arc. We shall compute the far field with two different angular sampling spacings: (i) $\Delta\phi=1°$ resulting in $N_\phi=360$, and (ii) $\Delta\phi=2°$ resulting in $N_\phi=180$. Hence, the relevant taper functions are given by the summation in Eq. (37) truncated at $|m|=180$ and $|m|=90$. Figure 11 shows upper envelopes of these taper functions for piston probes with h

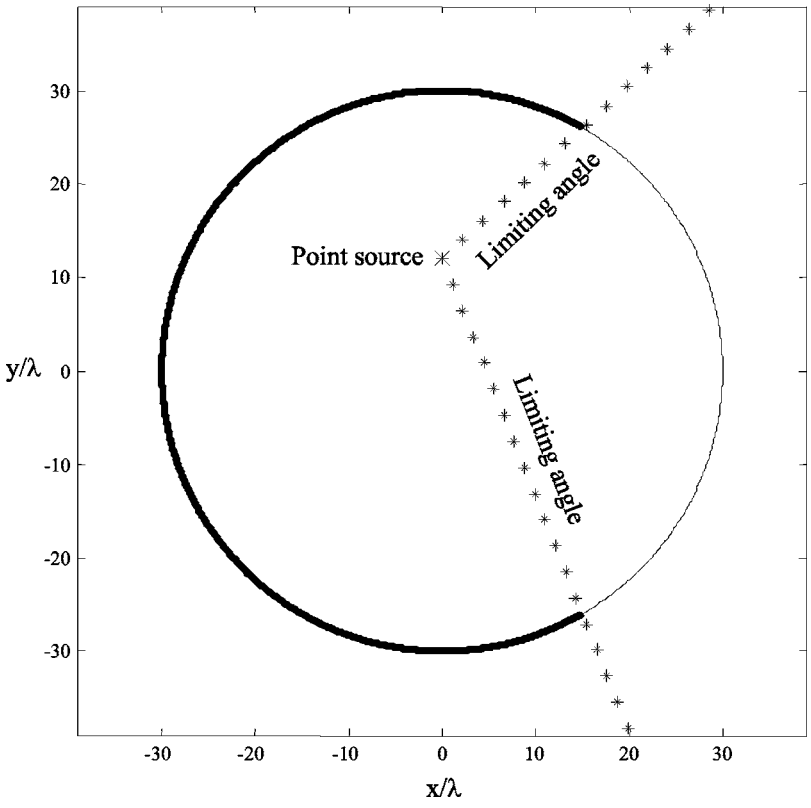


FIG. 10. Truncated scan cylinder with radius $a=30\lambda$. The scanned region is the solid circular arc given by $60°<\phi<300°$. The point source is located at $(x,y,z)=(0,12\lambda,-5\lambda)$, and the two limiting angles are marked by *.

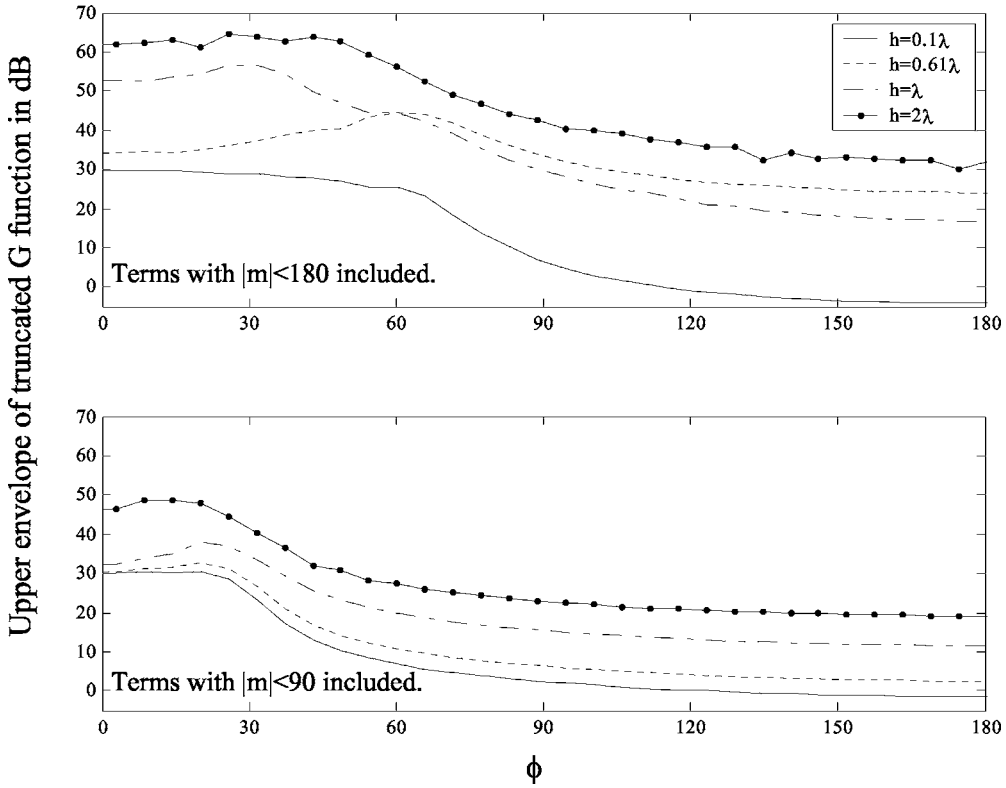


FIG. 11. Upper envelope of the taper functions of piston probes with $h=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$ and $\theta=90°$. Terms with $|m|<180=0.95ka$ are included in the top plot. Terms with $|m|<90=0.48ka$ are included in the bottom plot.

$=0.1\lambda$, $h=0.61\lambda$, $h=\lambda$, and $h=2\lambda$. The upper envelopes equal the maximum magnitude of the taper functions over 5° intervals.

These two sets of truncated taper functions are very different. For example, for $h=0.61\lambda$ the taper function truncated at $|m|=180$ (Fig. 11 top) equals 34 dB at $\phi=0$, attains a maximum value of 47 dB at $\phi-60°$, and decays to 24 dB at $\phi=180°$. The corresponding taper function truncated at $|m|=90$ (Fig. 11 bottom) equals 31 dB at $\phi=0$, attains a maximum value of 40 dB at $\phi=20°$, and decays to 3 dB at $\phi=180°$. These two truncated taper functions for the piston with $h=0.61\lambda$ lead to very different ϕ-truncation errors, as we now shall see.

Figure 12 shows the upper envelope of the far-field error

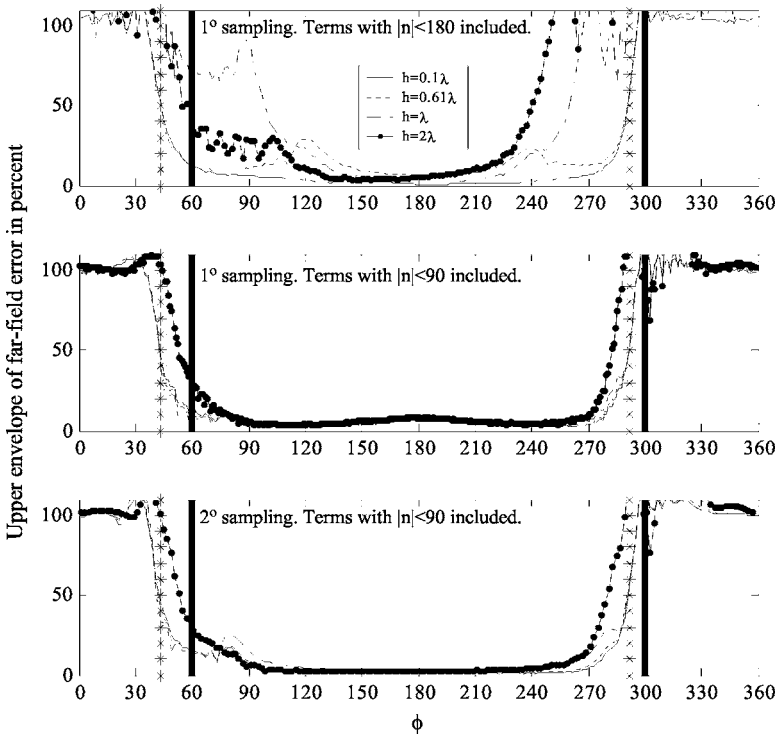


FIG. 12. Upper envelope for the error of the far field at $\theta=90°$ computed from a $\phi-z$ truncated scan with $a=30\lambda$, $60°<\phi<300°$, and $-40\lambda<z<39.5\lambda$. The point source is located at $(x,y,z)=(0,12\lambda,-5\lambda)$ and the angular boundaries of the scan are indicated with vertical solid lines and the limiting angles are marked by *. Top plot: angular sampling is 1° and terms with $|m|<180$ are included. Middle plot: angular sampling is 1° and terms with $|m|<90$ are included. Bottom plot: angular sampling is 2° and terms with $|m|<90$ are included.

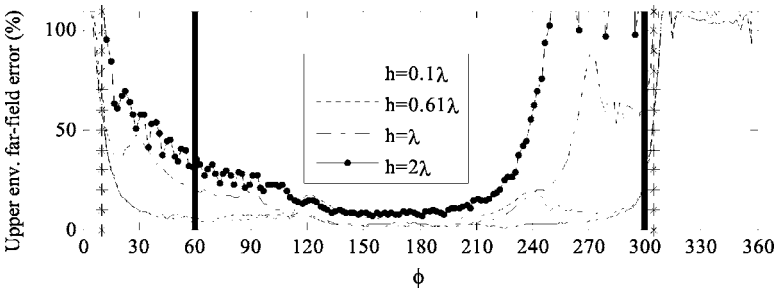


FIG. 13. Upper envelope for the error of the far field at $\theta=90°$ computed from a $\phi-z$ truncated scan with $a=30\lambda$, $60°<\phi<300°$, and $-40\lambda<z<39.5\lambda$. The point source is located at $(x,y,z)=(-17\lambda,20\lambda,-5\lambda)$ and the angular boundaries of the scan are indicated with vertical solid lines and the limiting angles are marked by *. The angular sampling is 1° and terms with $|m|<180$ are included.

computed from the truncated far-field summation (45) with the expansion coefficients determined from Eq. (48) using three different sets of parameters.

- Top plot: $\Delta\phi=1°$ and $N_0=180$ used in Eq. (45).
- Middle plot: $\Delta\phi=1°$ and $N_0=90$ used in Eq. (45).
- Bottom plot: $\Delta\phi=2°$ and $N_0=90$ used in Eq. (45).

Let us again consider the piston probe with $h=0.61\lambda$ for which the taper function in the top plot of Fig. 11 has a peak at 60°. Hence, for $\Delta\phi=1°$ and $N_0=180$ we should expect large errors when a truncation angle is 60° away from the observation direction. Indeed, Fig. 12 (top) shows that the far-field error has peaks at $\phi=120°$ and $\phi=240°$.

Next consider the middle plot where the sample spacing still is 1° but the far-field summation (45) is artificially truncated at $N_0=90$ (only these modes are required to compute the far field of the point source since $kR_{\min}=75$). Hence, neither of the taper functions in Fig. 11 apply. Except for a region near $\phi=180°$, the far-field error for this set of parameters is much smaller than in the top plot where all the modes obtained with 1° sampling are included.

The peaks at $\phi=80°$ and $\phi=280°$ for the probe with $h=0.61\lambda$ in the bottom plot of Fig. 12 can be explained from the taper function in the bottom plot of Fig. 11 that peaks at $\phi=20$.

Finally, Fig. 13 shows the far-field error for a point source at $\mathbf{r}_s=-17\lambda\hat{\mathbf{x}}+20\lambda\hat{\mathbf{y}}-5\lambda\hat{\mathbf{z}}$ computed with $\Delta\phi=1°$ and $N_0=180$ (for this point source $kR_{\min}=165$, so the terms with $|n|<90$ are not sufficient to compute the far field). The taper function in the top plot of Fig. 11 applies, and we confirm that the far-field error has peaks at $\phi=120°$ and $\phi=240°$ for $h=0.61\lambda$. Similar observations hold for the other piston sizes.

We have now demonstrated that the taper functions correctly predict the positions of the peaks in the far-field error. Moreover, the overall errors obtained with the truncated far-field formula (45) are larger for $N_0=180$ than for $N_0=90$. We can explain this from the behavior of the correction factor in Fig. 6 as follows. When only the terms with $|n|<90$ are included, only the two peaks at $n\simeq\pm60$ for the $h=2\lambda$ probe come into play. When all the terms with $|n|<180$ are included, all the peaks in Fig. 6 come into play except the two peaks at $n\simeq\pm190$ for the $h=0.61\lambda$ probe. As discussed in Sec. II B 3, the peaks of the correction factor cause inaccuracies in the computed far field for truncated scans, and therefore the overall far-field errors grow when more terms are included. The approximate formula (32) for the correction formula correctly predicts the locations of the peaks and is therefore useful for determining if a given sample spacing causes high far-field errors.

The far-field error depends therefore not only on the probe receiving characteristic, but also on the sample spacing used in the far-field computation and the number of terms that are included in the far-field computation. In practice, it is therefore advisable to plot the relevant correction factors and taper functions for different sample spacings to determine optimal probe and sampling parameters that reduce the far-field errors caused by ϕ truncation.

VI. CONCLUSIONS

Probe-corrected formulas for cylindrical near-field scanning of acoustic fields and the corresponding computation schemes were derived from outgoing and standing cylindrical wave functions. We developed edge-correction formulas that reduce the effect of z truncation and demonstrated their validity through numerical examples involving a point source and baffled piston transducer probes.

We showed that the probe-corrected formulas automatically apply a ϕ-dependent taper function to the probe output on the scan cylinder. The taper function depends on the probe and on the angular sample spacing. The taper function helps reduce the effect of angular truncation of the scan cylinder, and its shape determines the region of validity of the computed far field.

Far fields computed with small probes that have null-free patterns are insensitive to angular truncation and can be accurate in large angular regions. On the other hand, far-fields computed with larger probes that have nulls in their patterns are very sensitive to angular truncation and to the angular sample spacing employed during the measurement procedure. Therefore, it is advisable to use null-free probes. Regardless of the probe, the taper function can be used to optimize the parameters of the near-field scanning system to reduce the far-field errors caused by angular truncation.

ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research. The author would like to thank Arthur D. Yahjian for helpful discussions.

[1]D. M. Kerns, "Analytical techniques for the correction of near-field antenna measurements made with an arbitrary but known measuring antenna," *Abstracts of URSI-IRE Meeting*, Washington DC, April–May 1963, pp. 6–7.
[2]D. M. Kerns, "Correction of near-field antenna measurements made with

an arbitrary but known measuring antenna," Electron. Lett. **6**, 346–347 (1970).
[3]D. M. Kerns, "Scattering-matrix description and near-field measurements of electro-acoustic transducers," J. Acoust. Soc. Am. **57**, 497–507 (1975).
[4]D. M. Kerns, *Plane-Wave Scattering-Matrix Theory of Antennas and Antenna-Antenna Interactions* NBS Monograph 162 (US Government Printing Office, Washington D.C., 1981).
[5]T. B. Hansen and A. D. Yaghjian, *Plane-Wave Theory of Time-Domain Fields* (IEEE, New York, 1999).
[6]W. M. Leach and D. T. Paris, "Probe compensated near-field measurements on a cylinder," IEEE Trans. Antennas Propag. **AP-21**, 435–445 (1973).
[7]D. T. Paris, W. M. Leach, Jr., and E. B. Joy, "Basic Theory of probe-compensated near-field measurements," IEEE Trans. Antennas Propag. **AP-26**, 373–379 (1978).
[8]A. D. Yaghjian, *Near-Field Antenna Measurements on a Cylindrical Surface: a Source Scattering-Matrix Formulation*, NBS Technical Note 696 (revised) (National Bureau of Standards, Boulder, CO, 1977).
[9]G. V. Borgiotti, "An integral equation formulation for probe corrected far-field reconstruction from measurements on a cylinder," *IEEE Antenna Propagat. Soc. International Symposium Digest*, Stanford University, June 1977.
[10]G. V. Borgiotti, "Integral equation formulation for probe corrected far-field reconstruction from measurements on a cylinder," IEEE Trans. Antennas Propag. **26**, 572–578 (1978).
[11]F. Jensen, "Electromagnetic near-field far-field correlations," Ph.D. dissertation, Technical University of Denmark, Lyngby, Denmark, 1970.
[12]F. Jensen, "On the probe compensation for near-field measurements on a sphere," AEU **29**, 305–308 (1975).
[13]P. F. Wacker, "Near-field antenna measurements using a spherical scan: Efficient data reduction with probe correction," *IEE Conf. Publ. 113* (Conf. Precision Electromagn. Measurements, London, 1974), pp. 286–288.
[14]F. H. Larsen, "Probe Correction of spherical near-field measurements," Electron. Lett. **13**, 393–395 (1977).
[15]A. D. Yaghjian and R. C. Wittmann, "The receiving antenna as a linear differential operator: application to spherical near-field measurements," IEEE Trans. Antennas Propag. **AP-33**, 1175–1185 (1985).
[16]A. D. Yaghjian, "Simplified approach to probe-corrected spherical near-field scanning," Electron. Lett. **20**, 195–196 (1984).
[17]J. Hald, F. Jensen, and F. H. Larsen, *Spherical Near-Field Antenna Measurements*, edited by J. E. Hansen (Peter Peregrinus, London, 1988).
[18]R. C. Wittmann, "Probe-corrected spherical near-field scanning in acoustics," IEEE Trans. Instrum. Meas. **41**, 17–21 (1992).
[19]T. B. Hansen, "Formulation of spherical near-field scanning for time-domain acoustic fields," J. Acoust. Soc. Am. **98**, 1204–1215 (1995).
[20]T. B. Hansen, "Formulation of spherical near-field scanning for time-domain electromagnetic fields," IEEE Trans. Antennas Propag. **45**, 620–630 (1997).
[21]M. G. Cote and R. M. Wing, "Demonstration of bistatic electromagnetic scattering measurements by spherical near-field scanning," *Proc. AMTA Symposium* (Dallas, TX, 1963), p. 191.
[22]T. B. Hansen, R. A. Marr, U. H. W. Lammers, T. J. Tanigawa, and R. V. McGahan (unpublished).
[23]R. A. Marr, U. H. W. Lammers, T. B. Hansen, T. J. Tanigawa, and R. V. McGahan (unpublished).
[24]B. P. Hildebrand and B. B. Brenden, *An Introduction to Acoustical Holography* (Plenum Press, New York, 1974).
[25]E. G. Williams and J. D. Maynard, "Holographic Imaging without the Wavelength Resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[26]E. G. Williams, H. D. Hardy, and R. G. Fink, "Nearfield Acoustical Holography using an Underwater, Automated Scanner," J. Acoust. Soc. Am. **78**, 789–798 (1985).
[27]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, New York, 1999).
[28]J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), sec. 7.2.
[29]N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York, 1986), p. 220.
[30]A. D. Yaghjian, "Generalized or adjoint reciprocity relations for electroacoustic transducers," J. Res. Natl. Bur. Stand., Sect. B **79B**, 17–39 (1975).
[31]H. Weyl, "Ausbreitung elektromagnetischer Wellen uber einem ebenen Leiter," Ann. Phys. **60**, 481–500 (1919).
[32]L. V. King, Can. J. Res. **11**, 135–146 (1934).
[33]G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[34]E. B. Miller and A. D. Yaghjian, "Two theoretical results suggesting a method for calibrating ultrasonic transducers by measuring the total nearfield force," J. Acoust. Soc. Am. **66**, 1601–1608 (1979).
[35]H. Jeffreys, *Asymptotic Approximations* (Clarendon Press, Oxford, 1962), sec. 4.3.
[36]A. D. Yaghjian, *Upper-Bound Errors in Far-Field Antenna Parameters Determined from Planar Near-Field Measurements*, NBS Technical Note 667 (National Bureau of Standards, Boulder, CO, 1975).
[37]A. C. Newell, "Error analysis techniques for planar near-field measurements," IEEE Trans. Antennas Propag. **36**, 754–768 (1988).
[38]E. B. Joy and A. D. Dingsor, "Computer simulation of cylindrical surface near-field measurement system errors," *IEEE Antenna Propagat. Soc. International Symposium Digest*, Seattle, WA, 1979, pp. 565–568.
[39]F. Jensen, "Computer simulations as a design tool in near-field testing," *IEE Conf. Publ. 169, Pt. 1*, Seattle, WA, 1978, pp. 111–114.
[40]A. D. Yaghjian (private communication, 2002).
[41]D. S. Jones, *Acoustic and Electromagnetic Waves* (Clarendon Press, Oxford, 1986), pp. 417–421.
[42]A. D. Yaghjian, "Sampling criteria for resonant antennas and scatterers," J. Appl. Phys. **79**, 7474–7482 (1996); A. D. Yaghjian, Erratum **79**, 2547(E) (1996).

Obtaining the complex pressure field at the hologram surface for use in near-field acoustical holography when pressure and in-plane velocities are measured

Michael C. Harris and Jonathan D. Blotter[a)]
Department of Mechanical Engineering, Brigham Young University, 435 CTB, Provo, Utah 84602

Scott D. Sommerfeldt
Department of Physics and Astronomy, Brigham Young University, N281A ESC, Provo, Utah 84602

(Received 12 April 2005; revised 31 October 2005; accepted 28 November 2005)

Acoustical-based imaging techniques have found merit in determining the behavior of vibrating structures. These techniques are commonly used in numerous applications to obtain detailed noise source information and energy distributions on source surfaces. Source reconstructions using near-field acoustical holography (NAH) are reliant upon accurate measurement of the pressure field at the hologram surface. For complex acoustic fields this requires fine spatial resolution and therefore demands large microphone arrays. In this paper, an interpolation method is developed for obtaining the complex pressure field at the hologram surface from pressure and velocity measurements. Because particle velocity measurements provide directional information, a more accurate characterization of the pressure field with fewer measurement locations is obtained. The processing technique presented does not relate directly to the holographic reconstruction itself. However, the interpolation scheme presented serves as a preprocessing step before a NAH algorithm is applied. The presentation and validation of the interpolation scheme is the major focus of the paper. An analytical comparison of NAH reconstructions from traditional pressure measurements to reconstructions using the preprocessed pressure and velocity measurements is presented. A vibrating plate and cylinder are considered as test cases to validate the analytical results. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159427]

PACS number(s): 43.20.Ye, 43.40.At, 43.60.Pt [EGW] Pages: 808–816

I. INTRODUCTION

Near-field acoustic holography (NAH) is a methodology that enables the reconstruction of acoustic quantities in three-dimensional space from a two-dimensional measurement of the pressure field near the surface. Williams and Maynard presented a Fourier transform-based NAH method[1–3] for separable geometries of the wave equation that has been successfully applied to a variety of radiation problems.[4–6] Two approaches are currently available for arbitrary geometry problems. The first technique solves the Helmholtz integral equation numerically via the inverse boundary element method (IBEM).[7,8] An alternative to IBEM is the Helmholtz equation least-squares (HELS)[9,10] method which reconstructs the acoustic field using spherical basis functions.

One common aspect of all three NAH implementations is that the accuracy of reconstruction is dependent upon adequate representation of the pressure field on the measurement surface. The Fourier transform method and the IBEM rely on a spatial sampling for field characterization, which can cause mid- to high-frequency measurements to become cumbersome. This is due to the fact that the microphone spacing must be less than or equal to a half wavelength of the highest frequency of interest to avoid spatial aliasing.

The objective of this work is to develop a processing technique for pressure and velocity measurements that provides an improved characterization of the pressure field at the hologram surface. The result would be a reduction in the number of required measurement locations. This would lead to a considerable savings in data acquisition time for scanning array systems and help reduce the inefficiencies encountered at high frequencies.

Acoustic pressure and quadratic pressure measurements were used for NAH by Loyau *et al.*[11] in their development of broadband acoustic holography from intensity measurements (BAHIM) to obtain the phase of the pressure hologram without the need of a reference for scanning array systems. Visser[12] also showed that a particle velocity based adaptive boundary element method performs better than the conventional pressure based implementation if the pressure and velocity measurements have comparable signal-to-noise ratios.

This work relies on the ability to accurately measure acoustic particle velocity. Presently, the primary technique for particle velocity estimation is via finite difference approximations. The accuracy of this method depends on error in the pressure difference, scattering and diffraction, and microphone phase mismatch. Recently, a new particle velocity transducer known as a Microflown[13] sensor has been developed which functions similar to a hot wire anemometer. The transducer consists of two thin, parallel wires five microns apart that are heated to approximately 300 °C. As air particles flow across the wires heat transfer occurs. The first wire crossed will heat the air slightly which results in the second wire not being cooled to quite the same degree. This

[a)]Electronic mail: jblotter@byu.edu

temperature difference is then used to determine the particle velocity. Jacobsen and de Bree[14] showed that results comparable to finite difference intensity approximations are possible using the Microflown to measure the particle velocity. For the work presented in this paper, the Microflown sensor is used. However, the results are applicable to any sensor that measures both pressure and particle velocity, such as the energy density probe discussed in Ref. 15.

This paper presents a method where near-field pressure and in-plane velocities are used to obtain a representation of the pressure field at the measurement contour. The pressure and velocity field information are combined using a modified bicubic Hermite interpolation scheme presented in Sec. II. The work presented in this paper does not modify the currently used NAH algorithms discussed earlier. It does, however, provide the user with a better characterization of the field on the measurement surface to input into one of the currently available NAH algorithms. Analytical results are presented to indicate the theoretical benefits of the proposed preprocessing method. Experimental results for planar and cylindrical test cases are included for model validation. Because separable geometries have been chosen, the Fourier transform-based NAH method is implemented. However, the interpolation method presented in the following is applicable to arbitrary geometry problems if the IBEM or HELS method is selected.

II. PRESSURE FIELD INTERPOLATION THEORY

Current NAH reconstruction methods are based solely on measurement of the pressure field. Since pressure is a scalar quantity, it does not provide directional information for the field. Particle velocity measurements, on the other hand, supply first derivative information for the pressure field via Euler's equation,

$$\rho_0 \frac{\partial \vec{u}}{\partial t} = -\nabla p. \tag{1}$$

The measured in-plane velocities make derivative information available that is used to interpolate between measurement locations. This effectively simulates a finer mesh of pressure measurements.

Hermite Interpolation. The chosen interpolation method is taken from the area of geometric modeling.[16] For ease of programming and computability, along with other reasons specific to geometric modeling, the preferred way to perform interpolation is with parametric equations. For example, a three-dimensional curve is defined by $x=x(r,s)$, $y=y(r,s)$, and $z=z(r,s)$. It is generally convenient to normalize the domain of the parametric variables, r and s, by restricting their values to the closed interval between 0 and 1, inclusive. This restriction is expressed symbolically as $r,s \in [0,1]$. This interval establishes the bounding curves and the intermediate interpolation points. These curves have a natural vector representation given by

$$\mathbf{f}(r,s) = [x(r,s)\ y(r,s)\ z(r,s)]. \tag{2}$$

Farin[17] points out that a piecewise lower order polynomial interpolation approach is superior in speed and accuracy to its higher order counterparts. Therefore, bicubic polynomial interpolation is selected. Hermite surface patches are chosen for interpolation between measurement locations because they match both function values and slopes at the specified corner points.

Curves. Bicubic Hermite surfaces are composed of an orthogonal net of cubic Hermite curves. Therefore, a preliminary discussion of these curves is necessary to provide the foundation upon which the surface interpolation is built. For a more detailed development of Hermite interpolation see Ref. 16. The algebraic form of a parametric cubic curve is given by the polynomials in the following:

$$\begin{aligned} x(r) &= a_x r^3 + b_x r^2 + c_x r + d_x, \\ y(r) &= a_y r^3 + b_y r^2 + c_y r + d_y, \\ z(r) &= a_z r^3 + b_z r^2 + c_z r + d_z. \end{aligned} \tag{3}$$

The 12 scalar coefficients, called algebraic coefficients, determine a unique curve. Using vector notation to obtain a more compact form, Eq. (3) becomes

$$\mathbf{f}(r) = \mathbf{a}r^3 + \mathbf{b}r^2 + \mathbf{c}r + \mathbf{d}, \tag{4}$$

where $\mathbf{f}(r)$ is the position vector of any point on the curve and $\mathbf{a},\mathbf{b},\mathbf{c}$, and $\mathbf{d}$ are the vector equivalents of the scalar algebraic coefficients. The algebraic coefficients are not the most convenient way of controlling the shape of a curve, nor do they provide an intuitive sense of the curve shape. Converting to the Hermite form allows for the definition of conditions at the curve boundaries, or end points. Using the end points $\mathbf{f}(0)$ and $\mathbf{f}(1)$, the corresponding tangent vectors $\mathbf{f}^r(0)$ and $\mathbf{f}^r(1)$, and Eq. (4) yields the

$$\begin{aligned} \mathbf{f}(0) &= \mathbf{d}, \\ \mathbf{f}(1) &= \mathbf{a} + \mathbf{b} + \mathbf{c} + \mathbf{d}, \\ \mathbf{f}^r(0) &= \mathbf{c}, \\ \mathbf{f}^r(1) &= 3\mathbf{a} + 2\mathbf{b} + \mathbf{c}, \end{aligned} \tag{5}$$

where substituting $r=0$ into Eq. (4) yields $\mathbf{f}(0)$, and substituting $r=1$ into the equation yields $\mathbf{f}(1)$. Differentiating $\mathbf{f}(r)$ with respect to r obtains $\mathbf{f}^r(r)=3\mathbf{a}r^2+2\mathbf{b}r+\mathbf{c}$. Substituting $r=0$ and $r=1$ into this yields $\mathbf{f}^r(0)$ and $\mathbf{f}^r(1)$, respectively, where the superscript r indicates the derivative with respect to r. Solving this set of four simultaneous vector equations in four unknown vectors yields the algebraic coefficients in terms of the boundary conditions,

$$\begin{aligned} \mathbf{a} &= 2\mathbf{f}(0) - 2\mathbf{f}(1) + \mathbf{f}^r(0) + \mathbf{f}^r(1), \\ \mathbf{b} &= -3\mathbf{f}(0) + 3\mathbf{f}(1) - 2\mathbf{f}^r(0) - \mathbf{f}^r(1), \\ \mathbf{c} &= \mathbf{f}^r(0), \\ \mathbf{d} &= \mathbf{f}(0). \end{aligned} \tag{6}$$

Substituting these equations for the algebraic coefficient vectors into Eq. (4) and rearranging terms produces

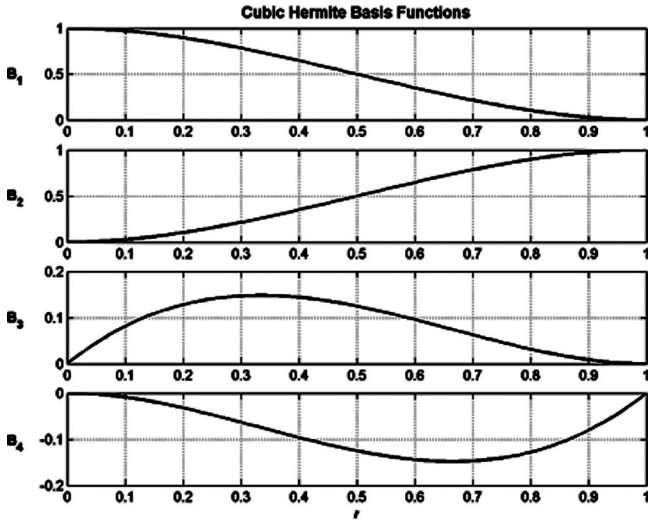


FIG. 1. Plot of the cubic Hermite basis functions in parameter space.

$$\mathbf{f}(r) = (2r^3 - 3r^2 + 1)\mathbf{f}(0) + (-2r^3 + 3r^2)\mathbf{f}(1) + (r^3 - 2r^2 + r)\mathbf{f}^r(0) + (r^3 - r^2)\mathbf{f}^r(1). \quad (7)$$

This equation is simplified by making the following substitutions:

$$\begin{aligned} \mathbf{B}_1(r) &= 2r^3 - 3r^2 + 1, \\ \mathbf{B}_2(r) &= -2r^3 + 3r^2, \\ \mathbf{B}_3(r) &= r^3 - 2r^2 + r, \\ \mathbf{B}_4(r) &= r^3 - r^2. \end{aligned} \quad (8)$$

Using these simplifications and subscripts to represent the end point r values, Eq. (7) becomes

$$\mathbf{f}(r) = \mathbf{B}_1(r)\mathbf{f}_0 + \mathbf{B}_2(r)\mathbf{f}_1 + \mathbf{B}_3(r)\mathbf{f}_0^r + \mathbf{B}_4(r)\mathbf{f}_1^r. \quad (9)$$

Equation (9) is called the geometric form, and the vectors $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_0^r$, and $\mathbf{f}_1^r$ are the geometric coefficients. The $\mathbf{B}_i(r)$ terms are called the Hermite basis functions. Figure 1 shows each basis function as a curve over the domain of the parameter r. These basis functions have three important characteristics. First, they are universal for all cubic Hermite curves. Second, they are only dependent on the parameter, making them identical for each of the three real space coordinates. Finally, they allow the constituent boundary condition coefficients to be decoupled from each other. These functions blend the effects of the end points and tangent vectors to produce the intermediate point coordinate values over the parameter domain. Letting

$$\mathbf{R} = [r^3 \quad r^2 \quad r \quad 1],$$

$$\mathbf{M}_H = \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad (10)$$

$$\mathbf{G}_H = [\mathbf{f}_0 \quad \mathbf{f}_1 \quad \mathbf{f}_0^r \quad \mathbf{f}_1^r]^T,$$

the geometric form given in Eq. (9) can be transformed into the more computationally efficient matrix form, where $\mathbf{M}_H$ is the Hermite basis transformation matrix and $\mathbf{G}_H$ is the geometry matrix containing the curve boundary conditions,

$$\mathbf{f}(r) = \mathbf{R}\mathbf{M}_H\mathbf{G}_H. \quad (11)$$

The geometry matrix in Eq. (11) is altered for each segment to obtain a series of cubic Hermite curves which are combined to form a composite curve with slope continuity at the end points.

Surfaces. A large complex surface can be defined by a composite collection of simpler patches. The algebraic form of a bicubic Hermite patch is given by the tensor product shown in the following:

$$\mathbf{f}(r,s) = \sum_{i=0}^{3}\sum_{j=0}^{3} \mathbf{a}_{ij} r^i s^j. \quad (12)$$

The $\mathbf{a}_{ij}$ are the three component algebraic coefficient vectors of the patch, where each component represents one of the three dimensions in real space. The subscripting corresponds to the order of the parameter variables that the coefficient is attributed to. Expanding Eq. (12) and arranging the $\mathbf{a}_{ij}$ terms in descending order produces Eq. (13), a 16 term polynomial in r and s,

$$\begin{aligned} \mathbf{f}(r,s) = {} & \mathbf{a}_{33}r^3s^3 + \mathbf{a}_{32}r^3s^2 + \mathbf{a}_{31}r^3s + \mathbf{a}_{30}r^3 + \mathbf{a}_{23}r^2s^3 \\ & + \mathbf{a}_{22}r^2s^2 + \mathbf{a}_{21}r^2s + \mathbf{a}_{20}r^2 + \mathbf{a}_{13}rs^3 + \mathbf{a}_{12}rs^2 \\ & + \mathbf{a}_{11}rs + \mathbf{a}_{10}r + \mathbf{a}_{03}s^3 + \mathbf{a}_{02}s^2 + \mathbf{a}_{01}s + \mathbf{a}_{00}. \end{aligned} \quad (13)$$

Because each of the 16 vector coefficients $\mathbf{a}_{ij}$ has three independent components, there are a total of 48 algebraic coefficients, or 48 degrees of freedom. In matrix notation, the algebraic form is

$$\mathbf{f}(r,s) = \mathbf{R}\mathbf{A}\mathbf{S}^T, \quad (14)$$

where

$$\mathbf{R} = [r^3 \quad r^2 \quad r \quad 1],$$

$$\mathbf{S} = [s^3 \quad s^2 \quad s \quad 1], \quad (15)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_{33} & \mathbf{a}_{32} & \mathbf{a}_{31} & \mathbf{a}_{30} \\ \mathbf{a}_{23} & \mathbf{a}_{22} & \mathbf{a}_{21} & \mathbf{a}_{20} \\ \mathbf{a}_{13} & \mathbf{a}_{12} & \mathbf{a}_{11} & \mathbf{a}_{10} \\ \mathbf{a}_{03} & \mathbf{a}_{02} & \mathbf{a}_{01} & \mathbf{a}_{00} \end{bmatrix}.$$

Since the $\mathbf{a}$ elements are three-component vectors, the $\mathbf{A}$ matrix is actually a $4\times4\times3$ array. As was found with Hermite curves, the algebraic coefficients of a Hermite patch determine its shape and position in space. Although the r,s parameter domain values are restricted between 0 and 1, the range of the variables in x,y, and z is not restricted, because the range of the algebraic coefficients is not limited. A unique point on the surface patch is generated each time a specific pair of r,s values are input into Eq. (14). These pairs of r,s values are then mapped back into real space.

Each patch is bounded by four curves, and each boundary curve is a cubic Hermite curve. Applying the same subscripting notation as implemented in Eq. (9), these curves are denoted as: $\mathbf{f}_{0s},\mathbf{f}_{1s},\mathbf{f}_{r0}$, and $\mathbf{f}_{r1}$, because they arise at the limit values of the parametric variables. There are also four unique corner points $\mathbf{f}_{00},\mathbf{f}_{01},\mathbf{f}_{10}$, and $\mathbf{f}_{11}$. As was seen for curves, the geometric form is a more convenient and intuitive way to define a patch. The geometric form is derived in the same way as for curves. The boundary conditions of the patch are used to solve for the algebraic coefficients. These conditions include the four patch corner points $\mathbf{f}_{00},\mathbf{f}_{01},\mathbf{f}_{10},\mathbf{f}_{11}$ and the eight tangent vectors $\mathbf{f}^r_{00},\mathbf{f}^s_{00},\mathbf{f}^r_{10},\mathbf{f}^s_{10},\mathbf{f}^r_{01},\mathbf{f}^s_{01},\mathbf{f}^r_{11},\mathbf{f}^s_{11}$ which define the boundary curves. $\mathbf{B}$ once again represents the Hermite basis functions, as in Eq. (9),

$$\begin{aligned}
\mathbf{f}(r,0) &= \mathbf{B}(r)[\mathbf{f}_{00} \quad \mathbf{f}_{10} \quad \mathbf{f}^r_{00} \quad \mathbf{f}^r_{10}]^T,\\
\mathbf{f}(r,1) &= \mathbf{B}(r)[\mathbf{f}_{01} \quad \mathbf{f}_{11} \quad \mathbf{f}^r_{01} \quad \mathbf{f}^r_{11}]^T,\\
\mathbf{f}(0,s) &= \mathbf{B}(s)[\mathbf{f}_{00} \quad \mathbf{f}_{01} \quad \mathbf{f}^s_{00} \quad \mathbf{f}^s_{01}]^T,\\
\mathbf{f}(1,s) &= \mathbf{B}(s)[\mathbf{f}_{10} \quad \mathbf{f}_{11} \quad \mathbf{f}^s_{10} \quad \mathbf{f}^s_{11}]^T.
\end{aligned} \tag{16}$$

These four curves provide 12 of the 16 vectors needed to specify the 48 degrees of freedom. Four additional vectors at the corner points, called twist vectors, are used to fully specify the patch. Mathematically, these vectors are defined as follows:

$$\begin{aligned}
\mathbf{f}^{rs}_{00} &= \frac{\partial^2\mathbf{f}(r,s)}{\partial r\,\partial s} \quad \text{at } r=0, s=0,\\
\mathbf{f}^{rs}_{10} &= \frac{\partial^2\mathbf{f}(r,s)}{\partial r\,\partial s} \quad \text{at } r=1, s=0,\\
\mathbf{f}^{rs}_{01} &= \frac{\partial^2\mathbf{f}(r,s)}{\partial r\,\partial s} \quad \text{at } r=0, s=1,\\
\mathbf{f}^{rs}_{11} &= \frac{\partial^2\mathbf{f}(r,s)}{\partial r\,\partial s} \quad \text{at } r=1, s=1.
\end{aligned} \tag{17}$$

Calculating the mixed partial derivative of Eq. (13) yields

$$\begin{aligned}
\frac{\partial^2\mathbf{f}(r,s)}{\partial r\,\partial s} &= 9\mathbf{a}_{33}r^2s^2 + 6\mathbf{a}_{32}r^2s + 3\mathbf{a}_{31}r^2 + 6\mathbf{a}_{23}rs^2 + 4\mathbf{a}_{22}rs\\
&\quad + 2\mathbf{a}_{21}r + 3\mathbf{a}_{13}s^2 + 2\mathbf{a}_{12}s + \mathbf{a}_{11}.
\end{aligned} \tag{18}$$

Evaluating Eq. (18) at the corner points obtains

$$\begin{aligned}
\mathbf{f}^{rs}_{00} &= \mathbf{a}_{11},\\
\mathbf{f}^{rs}_{10} &= 3\mathbf{a}_{31} + 2\mathbf{a}_{21} + \mathbf{a}_{11},\\
\mathbf{f}^{rs}_{01} &= 3\mathbf{a}_{13} + 2\mathbf{a}_{12} + \mathbf{a}_{11},\\
\mathbf{f}^{rs}_{11} &= 9\mathbf{a}_{33} + 6\mathbf{a}_{32} + 3\mathbf{a}_{31} + 6\mathbf{a}_{23} + 4\mathbf{a}_{22} + 2\mathbf{a}_{21} + 3\mathbf{a}_{13}\\
&\quad + 2\mathbf{a}_{12} + \mathbf{a}_{11}.
\end{aligned} \tag{19}$$

Doing the same for the remaining 12 vectors provides the remaining 12 equations required to solve for the algebraic coefficients,

$$\begin{aligned}
\mathbf{f}_{00} &= \mathbf{a}_{00},\\
\mathbf{f}_{10} &= \mathbf{a}_{30} + \mathbf{a}_{20} + \mathbf{a}_{10} + \mathbf{a}_{00},\\
\mathbf{f}_{01} &= \mathbf{a}_{03} + \mathbf{a}_{02} + \mathbf{a}_{01} + \mathbf{a}_{00},\\
\mathbf{f}_{11} &= \mathbf{a}_{33} + \mathbf{a}_{32} + \mathbf{a}_{31} + \mathbf{a}_{30} + \mathbf{a}_{23} + \mathbf{a}_{22} + \mathbf{a}_{21} + \mathbf{a}_{20} + \mathbf{a}_{13}\\
&\quad + \mathbf{a}_{12} + \mathbf{a}_{11} + \mathbf{a}_{10} + \mathbf{a}_{03} + \mathbf{a}_{02} + \mathbf{a}_{01} + \mathbf{a}_{00},\\
\mathbf{f}^r_{00} &= \mathbf{a}_{10},\\
\mathbf{f}^s_{00} &= \mathbf{a}_{01},\\
\mathbf{f}^r_{10} &= 3\mathbf{a}_{30} + 2\mathbf{a}_{20} + \mathbf{a}_{10},\\
\mathbf{f}^s_{10} &= \mathbf{a}_{31} + \mathbf{a}_{21} + \mathbf{a}_{11} + \mathbf{a}_{01},\\
\mathbf{f}^r_{01} &= \mathbf{a}_{13} + \mathbf{a}_{12} + \mathbf{a}_{11} + \mathbf{a}_{10},\\
\mathbf{f}^s_{01} &= 3\mathbf{a}_{03} + 2\mathbf{a}_{02} + \mathbf{a}_{01},\\
\mathbf{f}^r_{11} &= 3\mathbf{a}_{33} + 3\mathbf{a}_{32} + 3\mathbf{a}_{31} + \mathbf{a}_{30} + 2\mathbf{a}_{23} + 2\mathbf{a}_{22} + 2\mathbf{a}_{21} + 2\mathbf{a}_{20}\\
&\quad + \mathbf{a}_{13} + \mathbf{a}_{12} + \mathbf{a}_{11} + \mathbf{a}_{10},\\
\mathbf{f}^s_{11} &= 3\mathbf{a}_{33} + 2\mathbf{a}_{32} + \mathbf{a}_{31} + 3\mathbf{a}_{23} + 2\mathbf{a}_{22} + \mathbf{a}_{21} + 3\mathbf{a}_{13} + 2\mathbf{a}_{12}\\
&\quad + \mathbf{a}_{11} + 3\mathbf{a}_{03} + 2\mathbf{a}_{02} + \mathbf{a}_{01}.
\end{aligned} \tag{20}$$

Solving this set of 16 simultaneous equations from Eqs. (19) and (20) for the algebraic coefficients in terms of the geometric inputs and rearranging terms yields

$$\begin{aligned}
\mathbf{f}(r,s) &= [\mathbf{B}_1(r) \quad \mathbf{B}_2(r) \quad \mathbf{B}_3(r) \quad \mathbf{B}_4(r)]\mathbf{G}_H\\
&\quad \times [\mathbf{B}_1(s) \quad \mathbf{B}_2(s) \quad \mathbf{B}_3(s) \quad \mathbf{B}_4(s)]^T,
\end{aligned} \tag{21}$$

where $\mathbf{G}_H$ is the Hermite geometry matrix shown in the following:

$$\mathbf{G}_H = \begin{bmatrix}
\mathbf{f}_{00} & \mathbf{f}_{01} & \mathbf{f}^s_{00} & \mathbf{f}^s_{01}\\
\mathbf{f}_{10} & \mathbf{f}_{11} & \mathbf{f}^s_{10} & \mathbf{f}^s_{11}\\
\mathbf{f}^r_{00} & \mathbf{f}^r_{01} & \mathbf{f}^{rs}_{00} & \mathbf{f}^{rs}_{01}\\
\mathbf{f}^r_{10} & \mathbf{f}^r_{11} & \mathbf{f}^{rs}_{10} & \mathbf{f}^{rs}_{11}
\end{bmatrix}. \tag{22}$$

Recalling from Eq. (11) that $\mathbf{B}(r)$ may be expressed as $\mathbf{RM}_H$, Eq. (21) may be further simplified to obtain the conventional geometric form given by

$$\mathbf{f}(r,s) = \mathbf{RM}_H\mathbf{G}_H\mathbf{M}_H\mathbf{S}^T. \tag{23}$$

The remaining intricacy of the interpolation relates to converting between real and parameter space. A simple method for mapping between the two domains is presented in the following. Figure 2 provides an example of a set of four corner points in x and y that could be used to define a patch. In this case, the r parameter corresponds to the x direction and the s parameter to the y direction. Each (x,y)

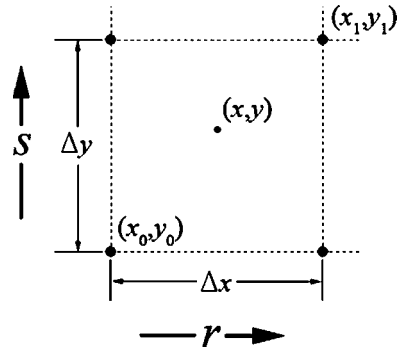


FIG. 2. Four corner points that could be used to define a Hermite surface patch.

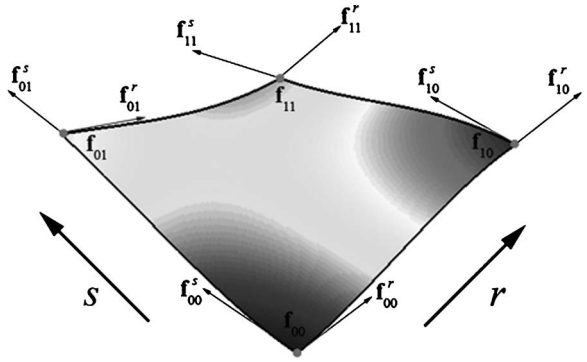


FIG. 3. A sample bicubic Hermite surface patch showing the required inputs at each corner point.

coordinate pair inside the patch corresponds to an (r,s) parameter pair. This parameter pair is obtained using

$$r = \frac{x - x_0}{\Delta x}, \qquad s = \frac{y - y_0}{\Delta y}, \tag{24}$$

where Δx and Δy correspond to the spacing between corner points in x and y, respectively. The slopes at the end points must also be transformed to the parameter domain. This is accomplished by scaling the r derivatives by Δx and the s derivatives by Δy as shown in Eq. (25) for the corner point corresponding to $r=s=0$,

$$\mathbf{f}^r_{00} = \frac{\partial \mathbf{f}_{00}}{\partial x} \cdot \frac{\Delta x}{\Delta r}, \qquad \mathbf{f}^s_{00} = \frac{\partial \mathbf{f}_{00}}{\partial y} \cdot \frac{\Delta y}{\Delta s}, \tag{25}$$

where Δr and Δs equal one because they are restricted to vary from zero to one. The form is the same for the remaining three corner points of the patch.

Because measurements with pressure and velocity sensors do not in general provide enough information to calculate twist vectors, they have been set to zero for this investigation. The Hermite geometry matrix from Eq. (22) then becomes

$$\mathbf{G}_H = \begin{bmatrix} \mathbf{f}_{00} & \mathbf{f}_{01} & \mathbf{f}^s_{00} & \mathbf{f}^s_{01} \\ \mathbf{f}_{10} & \mathbf{f}_{11} & \mathbf{f}^s_{10} & \mathbf{f}^s_{11} \\ \mathbf{f}^r_{00} & \mathbf{f}^r_{01} & 0 & 0 \\ \mathbf{f}^r_{10} & \mathbf{f}^r_{11} & 0 & 0 \end{bmatrix}. \tag{26}$$

This limits the patches to having only first derivative continuity at their edges. The results presented in the following indicate that adequate reconstructions are still obtained with this simplification. A second order fit could also provide first derivative continuity but lacks the ability to interpolate more than a half of a wavelength between measurements. Figure 3 shows a sample bicubic Hermite patch and the required inputs at each corner point $\mathbf{f}_{rs}$. Each patch represents the rectangular area between four corner point locations. The above-noted interpolation is repeated for each segment of the surface and all the patches combined.

III. ANALYTICAL IMPLEMENTATION

With the surface interpolation completed, the chosen NAH algorithm is applied. In this section an analytical model is developed to investigate the theoretical benefits of the pressure field preprocessing method. This model requires first that a synthetic acoustic field be created from a hypothetical source. The field is then sampled and the chosen algorithm implemented. The error is then evaluated on the estimation plane by comparing the actual and reconstructed fields.

A. Synthetic field creation

A rectangular, simply supported plate is chosen as the hypothetical source because it has a simple closed-form radiation equation. The plate shown in Fig. 4 is driven by a

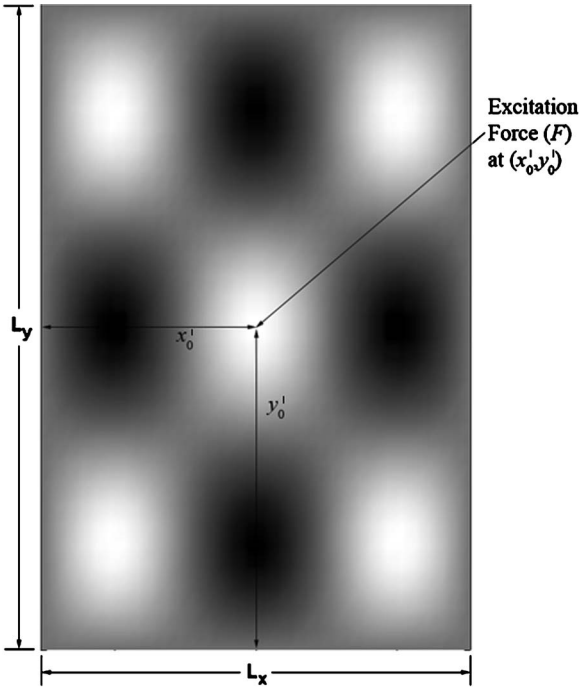


FIG. 4. Geometry of the simply supported plate chosen to be the source for the analytical modeling.

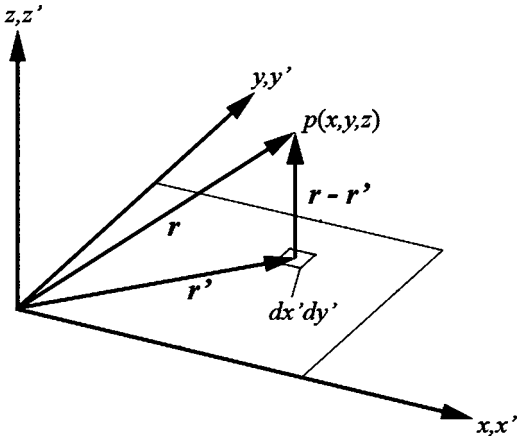


FIG. 5. Description of the geometric quantities used in Rayleigh's integral.

harmonic point source acting normal to the plate at its center (x_0', y_0'). The surface displacement w for the plate as a function of angular frequency ω is given by Eq. (27),[3]

$$w(x',y',\omega) = -\frac{F}{\rho h}\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}\frac{\Phi_{mn}(x_0',y_0')\Phi_{mn}(x',y')}{\omega^2-\omega_{mn}^2}, \quad (27)$$

$$\Phi_{mn}(x',y') = \frac{2}{\sqrt{L_xL_y}}\sin\left(\frac{m\pi x'}{L_x}\right)\sin\left(\frac{n\pi y'}{L_y}\right), \quad (28)$$

where F is the excitation force amplitude, ρ the plate material density per unit area, h the plate thickness, and L the length of the plate in the indicated direction. The subscripts m and n denote the plate mode numbers in the x and y directions, respectively. Assuming that the plate is in an infinite rigid baffle, the radiated pressure can be expressed in terms of the plate surface displacement using Rayleigh's integral. Figure 5 provides a clear description of the geometric quantities to be used in Eq. (29), where $e^{-i\omega t}$ time dependence has been assumed,

$$p(x,y,z,\omega) = -\frac{\omega^2\rho_0}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}w(x',y',\omega)\frac{e^{ik|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|}dx'dy'. \quad (29)$$

The pressure at a point in space, $p(x,y,z,\omega)$ is computed by summing the contribution from each $dx'dy'$ area element. Radiation from the plate is simulated using a discrete summation of Eq. (29) for a 32×32 grid of point sources on the plate. The field is then sampled at chosen measurement locations to obtain the pressure and gradient information to be used for interpolation. The selected NAH algorithm is then applied to reconstruct the field.

B. Error evaluation

The reconstruction error is evaluated by first calculating the pressure field at the measurement and estimation planes directly using Eq. (29). The direct calculation of the pressure field at the estimation plane serves as a reference against which the NAH reconstruction is compared. The reconstruction error is quantified by differencing the NAH estimation and the direct calculation at the estimation plane. The standard deviation of these residuals is then computed and normalized by the maximum pressure field value to obtain a single value representing the whole field error. The error for the reconstructions using the preprocessed data is compared to pressure only reconstructions. The number of sensors used to populate the measurement array is varied in both dimensions in order to determine the possible reduction in sensor count using the pressure and in-plane velocity measurements.

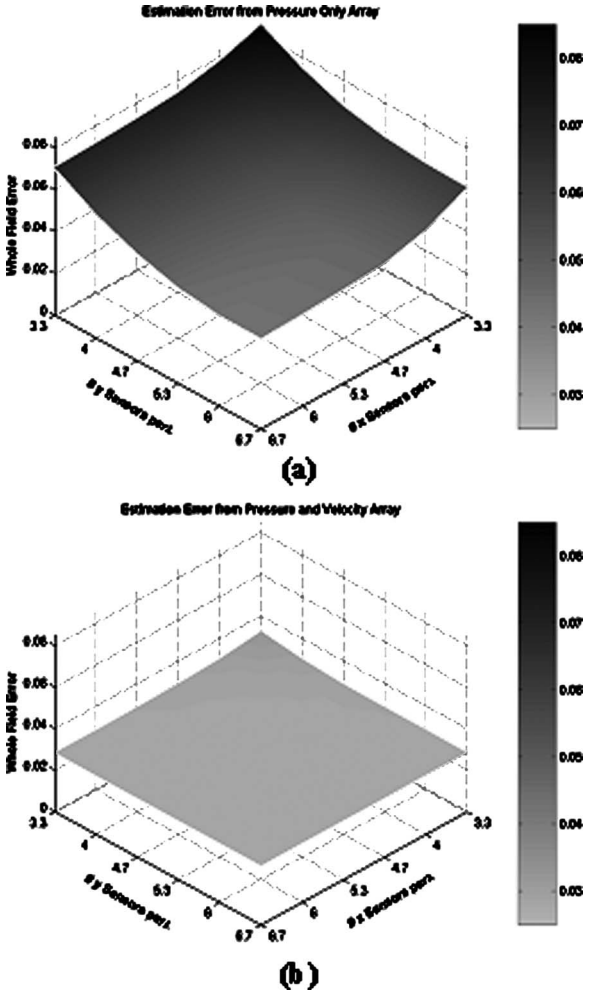


FIG. 6. (a) The estimation error for NAH reconstructions from pressure measurements as a function of sensors per spatial wavelength. (b) Estimation error for NAH reconstructions using preprocessed pressure and velocity measurements as a function of sensors per spatial wavelength.

C. Results

The results to follow correspond to the synthetic field generated by a 30.5 cm $\times$ 45.7 cm $\times$ 0.3175 cm plate vibrating in the 3,3 mode (1090 Hz), as shown in Fig. 4. These dimensions are chosen to match the dimensions used for the experimental validation presented in Sec. IV. The measurement plane is set to 5 cm and the estimation plane to 2 cm above the plate. Figures 6(a) and 6(b) show the resulting normalized whole field NAH estimation error plots for array sizes ranging from 10×10 to 20×20 for conventional pressure measurements and the processed pressure and velocity data. The data are presented as a function of the number of

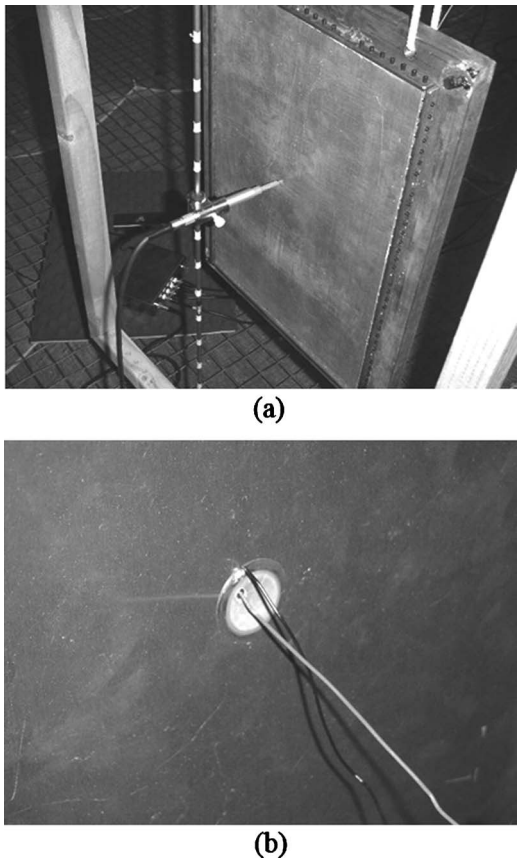


FIG. 7. (a) The experimental setup for the simply supported plate. (b) The piezoelectric patch used to excite the plate at its center.

sensors per spatial wavelength λ. This provides the reader with a qualitative sense of the relationship between estimation error and the sensor spacing.

These plots indicate that the inclusion of velocity field information at the hologram surface significantly improves the ability to reconstruct the field accurately. In fact, the NAH reconstruction using a 10×10 array of processed pressure and velocity measurements has a whole field error of 0.0326, which is slightly lower than the 0.0433 error for a 20×20 array of pressure measurements. These results show that the number of measurement locations can be reduced by

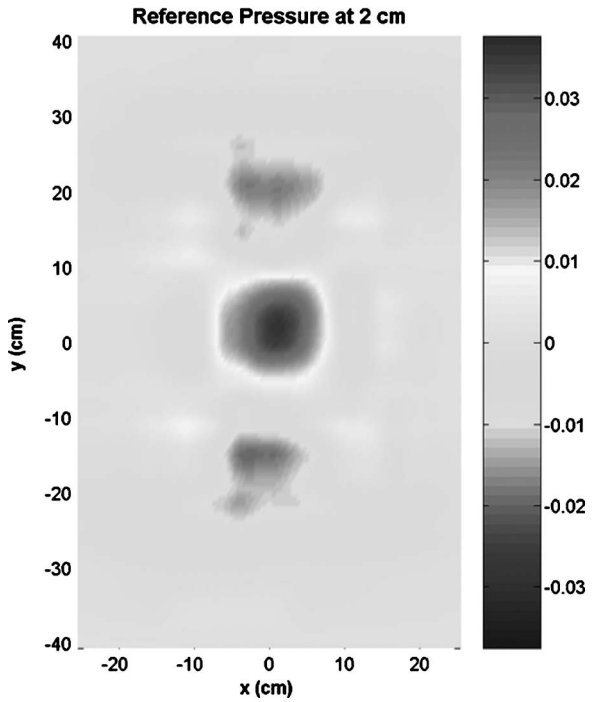


FIG. 8. The reference pressure at the 2 cm estimation plane against which the NAH reconstructions are compared.

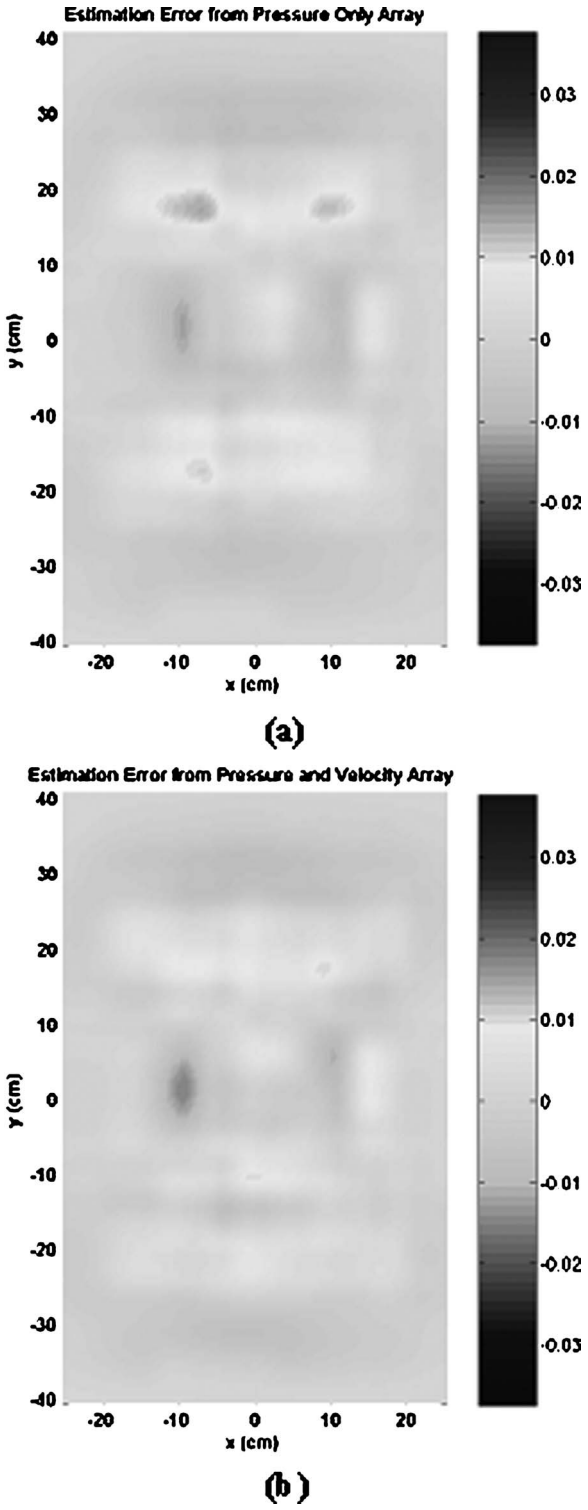


FIG. 9. (a) The pressure field reconstruction error at the 2 cm estimation plane from an 11×17 array of pressure measurements at 5 cm. (b) The pressure field reconstruction error at 2 cm from a 6×9 array of preprocessed pressure and in-plane velocity measurements at 5 cm.

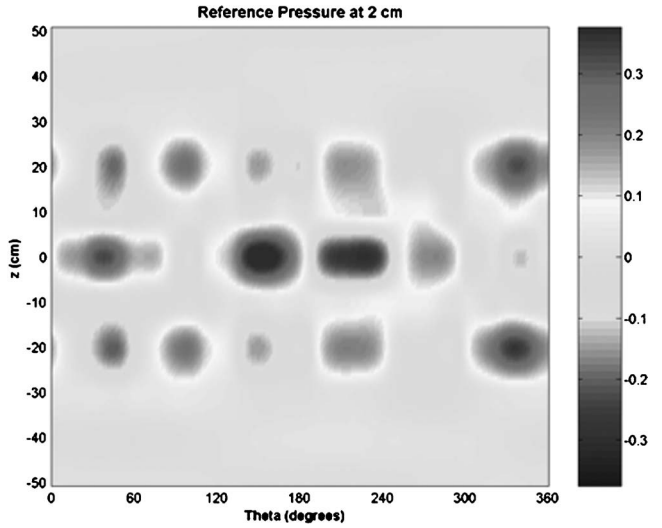


FIG. 10. The reference pressure field at 2 cm for the cylindrical test case.

about 75% when sensing equipment that measures pressure and velocity is used. This reduction seems reasonable since twice the information is being used in each direction. If a three channel probe is used to measure the field, a channel count reduction of 25% would also be realized for nonscanning systems. These results represent the theoretical optimal performance of the measurement schemes because the measurements have zero positioning, amplitude, and phase error.

IV. EXPERIMENTAL VALIDATION

A. Planar test case

An experimental setup is designed to approximate the simply supported plate used in the analytical investigation. Figure 7(a) shows the 30.5 cm $\times$ 45.7 cm $\times$ 0.3175 cm aluminum plate. It is attached along its edges to a heavy steel frame using cone point set screws to approximate the simply supported boundary condition. A 20-mm-diam piezoelectric patch, shown in Fig. 7(b), is used to excite the plate at its center. The plate and a measurement grid are suspended in an anechoic chamber for data acquisition. A single Microflown ultimate sound probe (USP) is used to scan the field to obtain the pressure and in-plane velocities required for the modified bicubic Hermite interpolation method. The plate is excited at 1090 Hz corresponding to the 3,3 operating shape. The field is sampled at 2 and 5 cm from the plate as in the analytical case. The vertical and horizontal step distance is set to 5 cm and the plate is overscanned in both directions yielding a 50 cm $\times$ 80 cm overall measurement array size. The 2 cm measurement again serves as the reference against which the NAH reconstructions are compared.

Figure 8 shows the reference pressure as measured on the 2 cm estimation plane. An 11 $\times$ 17 array of pressure measurements at 5 cm is used to reconstruct the pressure at the estimation plane using the traditional Fourier NAH method. The pressure only reconstruction is then compared to the reconstruction using a 6 $\times$ 9 array of pressure and velocity measurements spanning the same area. Figures 9(a) and 9(b) show the reconstruction error obtained by differencing the reference pressure and the reconstructions.

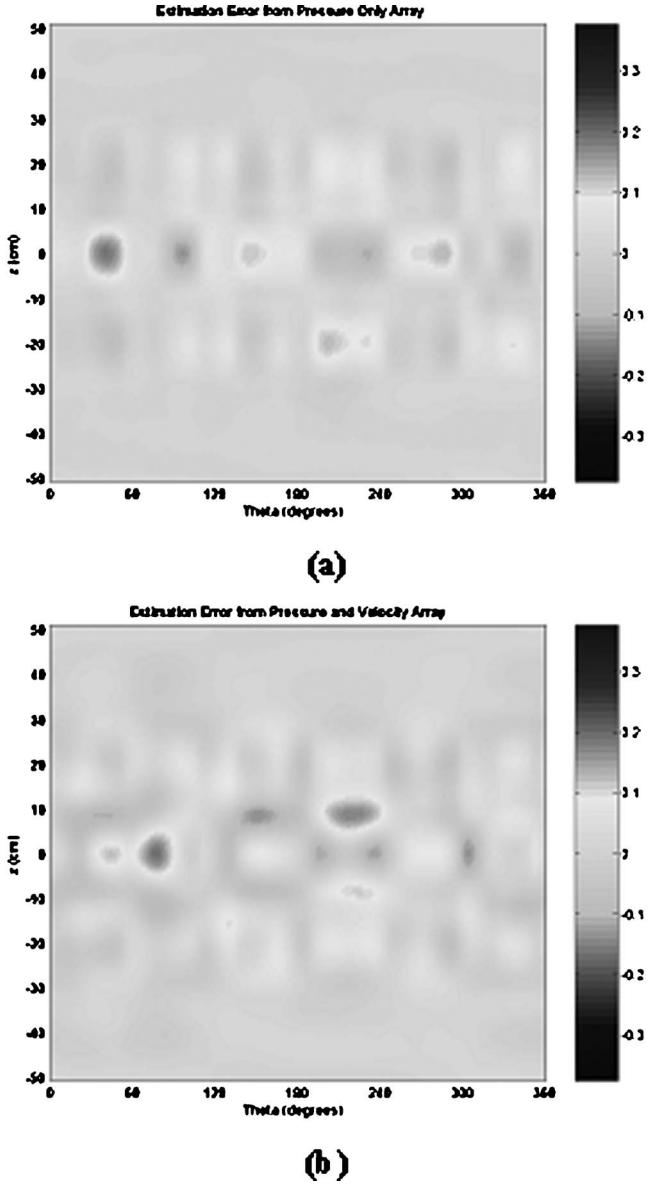


FIG. 11. (a) The pressure field reconstruction error at the 2 cm estimation plane from an 11 $\times$ 14 array of pressure measurements at 4 cm. (b) The pressure field reconstruction error at 2 cm from a 7 $\times$ 7 array of preprocessed pressure and in-plane velocity measurements at 4 cm.

Both NAH reconstructions are able to accurately characterize the pressure field on the estimation plane. The normalized whole field estimation error for the NAH reconstruction using only pressure measurements is 0.051, while the error for the reconstruction using the processed pressure and velocity data is 0.039. The reconstruction using the processed data is slightly more accurate than the conventional reconstruction with 70% fewer measurements. This corresponds well with the simulated results.

B. Cylindrical test case

A cylindrical ABS plastic tube is used for this test case. The tube dimensions are: 10.2 cm inner diameter, 10.8 cm outer diameter, 50.8 cm length. Simply supported boundary conditions are approximated at the tube ends using tapered conical plugs. The tube is driven at 1524 Hz with the same

20 mm piezoelectric patch used for the plate. This excitation corresponds to the 3,3 operating shape. Scans are made at 2 and 4 cm radial distances from the outer surface of the tube. Figure 10 shows the reference field at 2 cm. The resulting reconstruction error from an 11×14 array of pressure measurements at 4 cm is presented in Fig. 11(a). The vertical step distance is 10.2 cm and the incremental rotation angle is 27.7°. The reconstruction error shown in Fig. 11(b) is obtained using a 7×7 array of pressure and velocity measurements.

Both reconstructions retain the correct modal shape. However, the reconstruction utilizing the processed pressure and velocity measurements is more accurate at the peak amplitudes. The normalized whole field error for the conventional NAH reconstruction is 0.030. The error for the reconstruction based on pressure and velocity measurements is 0.024. Once again the NAH reconstruction using the proposed preprocessing method is able to provide slightly better reconstruction with 70% fewer measurements locations.

V. CONCLUSIONS

Based on the analytical and experimental results presented in this work, a 70% reduction in the number of measurement locations required for NAH is possible when pressure and in-plane velocities are used. For cases where subarrays of sensors are required to scan the field, the proposed preprocessing method reduces significantly the amount of repositioning, and therefore time, required. The subarrays could also be increased in size up to three and a half times if the same number of sensors is used. It should also be noted that the proposed interpolation method is applicable to other reconstruction methods, such as IBEM, that rely on a spatial sampling of the pressure field.

ACKNOWLEDGMENTS

The authors would like to thank Microflown Technologies for the use of the USP sensor. This work was supported by NASA and the National Science Foundation.

[1]E. G. Williams and J. D. Maynard, "Holographic imaging without the wavelength resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[2]J. D. Maynard, E. G. Williams, and Y. Lee, "Near-field acoustic holography. I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999).
[4]E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[5]T. Antoine, J. Catalfaud, and C. Aujard, "Meeting noise regulations with nearfield acoustic holography," Sound and Vibration Magazine, 18–20, May 2003.
[6]D. Vaucher de la Croix, P. Chevret, and F. Perrin, "Use of acoustical holography in 3D interiors measurements," in *Proceedings of International Congress and Exposition on Noise Control Engineering, Dearborn, MI, August 2002*.
[7]W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[8]G.-T. Kim and B.-H. Lee, "3-D sound source reconstruction and field reprediction using the helmholtz integral equation," J. Sound Vib. **136**, 245–261 (1990).
[9]Z. Wang and S. F. Wu, "Helmholtz equation-least squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[10]N. Rayess and S. F. Wu, "Experimental validations of the HELS method for reconstructing acoustic radiation from a complex vibrating structure," J. Acoust. Soc. Am. **107**, 2955–2964 (2000).
[11]T. Loyau, J. Pascal, and P. Gaillard, "Broadband acoustic holography reconstruction from acoustic intensity measurements. I. Principle of the method," J. Acoust. Soc. Am. **84**, 1744–1750 (1988).
[12]R. Visser, "Inverse source identification based on acoustic particle velocity measurements," in *Proceedings of the International Congress and Exposition on Noise Control Engineering*, INTERNOISE 2002, Dearborn, MI, August 2002.
[13]H. E. de Bree, "The Microflown: An acoustic particle velocity sensor," Acoust. Aust. **31**, 91–94 (2003).
[14]F. Jacobsen and H. E. de Bree, "A comparison of P-P and P-U sound intensity measurement systems," in *Proceedings of the Eleventh International Congress on Sound and Vibration*, St. Petersburg, Russia, July 2004, pp. 3159–3166.
[15]L. Locey, "Analysis and comparison of three acoustic energy density probes," Masters of Science Degree thesis, Brigham Young University, December 2004.
[16]M. Mortenson, *Geometric Modeling Second Edition* (Wiley, New York, 1997).
[17]G. Farin, *Curves and Surfaces for Computer-Aided Geometric Design* (Academic, New York, 1993).

Scattering by quasi-symmetric pipes

Michael Carley[a)]
Department of Mechanical Engineering, University of Bath, Bath BA2 7AY, England

(Received 3 June 2005; revised 7 October 2005; accepted 29 November 2005)

A hypersingular boundary integral method for the prediction of radiation from a straight circular pipe with arbitrary end profile has been developed. The technique represents an extension of established procedures for axisymmetric pipes with the addition of recent advances in special function and quadrature theory to simplify the implementation. The resulting code is applied to two sample problems: first, the prediction of radiation of a plane wave mode from a pipe with its ends cut by an inclined plane, representing the "scarfed intake" proposed for reduction of aircraft engine noise. Second, the method is used to examine scattering of an incident azimuthal mode by a multi-lobed profile, characteristic of the "chevron" nozzles proposed for jet engine exhausts.
 [DOI: 10.1121/1.2159432]

PACS number(s): 43.20.−f, 43.20.Mv, 43.20.Rz [MO] Pages: 817–823

I. INTRODUCTION

This paper addresses the problem of calculating the sound radiated from a straight, circular section pipe with an arbitrary end profile, motivated by two features proposed for modern aero-engines. The first, "scarfing," has been proposed as a means of redirecting the fan noise radiated from the intake of an engine so that the noise on the ground is reduced.[1] The second is the "chevron nozzle" in which an asymmetric exhaust nozzle is used to modify the flow so that the noise source is reduced.[2] In such an approach, the main reason for the noise reduction appears to be the modification of the flow and consequent effect on the aerodynamic source terms, but an examination of the purely acoustic effect of the chevron nozzle is still of some interest, if only for its connection with the instabilities of the jet shear layer.[3]

The method to be developed is a relatively conventional technique for scattering from straight circular pipes as used by a number of researchers in the past,[4–6] who employed a hypersingular integral equation to solve the scattering problem. The essential difference is that in this paper, the approach is extended to a pipe with end profiles given by an arbitrary function of azimuthal angle on the pipe. In addition, recent advances in quadrature theory[7] and special functions[8] are used to handle the hypersingular terms which arise in the integral equation. The central part of the method is the decomposition of the integral equation into a set of coupled equations for the amplitudes of the azimuthal modes of the surface pressure jump. This makes the problem into one similar to that in engine noise where a duct mode interacts with the internal stationary vanes in an engine[9] with the role of the stators being played by the duct end profile.

The problem of scattering by scarfed duct terminations has been studied with asymptotic methods being applied to the case of low-frequency radiation from ducts with small scarf angle[1] and ray theory to high-frequency radiation from ducts of arbitrary scarf angle.[10] There is then a gap in the literature which this paper aims to fill: the problem of relatively low-frequency scattering from duct terminations with large scarf angle. Since the general termination problem is no more difficult numerically than the simple scarfed duct problem, results for the chevron termination case will also be presented.

II. NUMERICAL TECHNIQUE

An integral equation method based on existing techniques is now developed for the solution of the problem of scattering by straight, circular pipes with arbitrary end profiles. The formulation yields a set of coupled equations for azimuthal mode amplitudes on a distorted circular source, not unlike the "wobbling modes" studied in propeller noise.[11,12] The resulting equations include a hypersingular integral which is evaluated using recently developed quadratures, easing the analysis of the integral kernel.

A. Formulation

The geometry of the system is shown in Fig. 1. A zero-thickness, circular section, straight pipe extending from $z=a$ to $z=b$ is subjected to an incident acoustic field p_{inc}. The pipe ends have identical azimuth-dependent profiles $\epsilon w(\theta)$ where ϵ is the profile depth and $w(\theta)$ is the shape; $-\sin\theta$ for a linear scarf. A cylindrical coordinate system (r,θ,z) is used for calculations with position on the pipe surface given by

$$x = r_1 \cos\theta_1, \quad y = r_1 \sin\theta_1, \quad z = z_1 + \epsilon w(\theta_1),$$

where the subscript 1 indicates variables of integration. The profile shape $\epsilon w(\theta_1)$ is imposed at all axial positions so that the formulation of the integral equations will involve analysis of radiation from distorted ring sources. For the directivity plots shown later, spherical coordinates (R,θ,ϕ) are used and are indicated in Fig. 1.

The total acoustic field is made up of the incident field and that radiated by the pipe surface. Using a standard approach,[4–6] the solution for the scattered field is computed via the pressure jump Δp across the surface, which acts as a distribution of radial dipoles:

[a)]Electronic mail: m.j.carley@bath.ac.uk

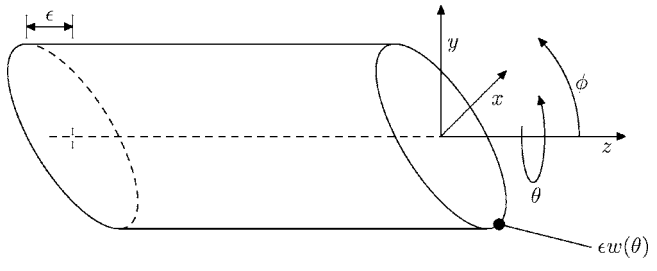


FIG. 1. Coordinate systems for scarfed cylinder.

$$p(r,\theta,z) = p_{\text{inc}} + \int\int \Delta p(\theta_1,z_1)\frac{\partial}{\partial r_1}\frac{e^{jkR}}{4\pi R}\,dS, \tag{1}$$

where S is the scattering surface, R is the source-observer distance, k is the acoustic wave number, and time dependence $\exp -j\omega t$ has been assumed.

To find Δp, the integral equation is solved subject to the boundary condition that the pressure gradient on the surface be identically zero:

$$-\frac{\partial p_{\text{inc}}}{\partial r} = \int\int \Delta p(\theta_1,z_1)\frac{\partial^2}{\partial r\,\partial r_1}\frac{e^{jkR}}{4\pi R}\,dS. \tag{2}$$

Because the two ends of the pipe have identical profiles, the transformation of the integral has unit Jacobian and Eq. (2) can be written

$$-\frac{\partial p_{\text{inc}}}{\partial r} = \int_a^b\int_0^{2\pi} \Delta p(\theta_1,z_1)\frac{\partial^2}{\partial r\,\partial r_1}\frac{e^{jkR}}{4\pi R}\,d\theta_1\,dz_1. \tag{3}$$

The standard approach[4–6] to solving the integral equation for Δp is to decompose it into a suitable set of basis functions and compute the coefficients of these functions subject to the boundary conditions. In this case, Δp is decomposed into azimuthal modes in $\exp jn\theta$ and Chebyshev polynomials of the second kind U_n so that

$$\Delta p(\theta_1,z_1) = \sum_{m=-\infty}^{\infty}\sum_{n=0}^{N} Q_{nm}e^{jm\theta_1}U_n\left(\frac{z_1-\bar{z}}{L/2}\right) \times (z_1-a)^{1/2}(b-z_1)^{1/2}, \tag{4}$$

where the pipe length $L=b-a$. Inserting Eq. (4) into Eq. (3), gives an integral equation for the coefficients Q_{nm}:

$$-\frac{\partial p_{\text{inc}}}{\partial r} = \sum_m Q_{nm}\sum_{n=0}^{N}\left(\frac{L}{2}\right)^2\int_{-1}^{1} U_n(t)(1-t^2)^{1/2}\frac{\partial^2 g_m}{\partial r\,\partial r_1}\,dt, \tag{5}$$

where

$$g_m = \int_0^{2\pi}\frac{e^{j(kR+m\theta_1)}}{4\pi R}\,d\theta_1,$$

$$\bar{z} = \frac{a+b}{2}, \quad z_1 = \bar{z} + \frac{L}{2}t.$$

Equation (5) is a set of coupled equations which must be solved for all azimuthal mode orders simultaneously, with the required mode orders m to be prescribed in advance.

B. Green's function analysis

To numerically evaluate the integral of Eq. (5), the singular part of the kernel must be isolated and analyzed. This is most easily handled using some recent results in the theory of elliptic-type integrals. Following Dawson,[13] the Green's function g_m is broken into a bounded and a singular part:

$$g_m = g_m^{(b)} + g_m^{(s)}, \tag{6a}$$

$$g_m^{(b)} = \int_0^{2\pi} e^{jm\theta_1}\frac{e^{jkR_\epsilon} - 1 - jkR_\epsilon + k^2R_\epsilon^2/2}{4\pi R_\epsilon}\,d\theta_1, \tag{6b}$$

$$g_m^{(s)} = \int_0^{2\pi} e^{jm\theta_1}\frac{1 + jkR_\epsilon - k^2R_\epsilon^2/2}{4\pi R_\epsilon}\,d\theta_1, \tag{6c}$$

$$R_\epsilon^2 = r^2 + r_1^2 - 2rr_1\cos(\theta-\theta_1) + (z - z_1 + \epsilon\Delta w)^2,$$

$$\Delta w = w(\theta) - w(\theta_1).$$

The kernel of Eq. (5) is found by differentiation so that the bounded part can be written

$$\begin{aligned}\frac{\partial^2 g_m^{(b)}}{\partial r\,\partial r_1} = \int_0^{2\pi}\frac{e^{jm\theta_1}}{4\pi}\Bigg[&\frac{2}{R_\epsilon^3}(e^{jkR_\epsilon} - 1 - jkR_\epsilon + k^2R_\epsilon^2/2) \\ &- \frac{jk}{R_\epsilon^2}(e^{jkR_\epsilon} - 1 - jkR_\epsilon) - \frac{k^2}{R_\epsilon}(e^{jkR_\epsilon} - 1)\Bigg]\frac{\partial R_\epsilon}{\partial r}\frac{\partial R_\epsilon}{\partial r_1} \\ &+ \frac{e^{jm\theta_1}}{4\pi}\Bigg[\frac{jk}{R_\epsilon}(e^{jkR_\epsilon} - 1 - jkR_\epsilon) \\ &- \frac{1}{R_\epsilon^2}(e^{jkR_\epsilon} - 1 - jkR_\epsilon + k^2R_\epsilon^2/2)\Bigg]\frac{\partial^2 R_\epsilon^2}{\partial r\,\partial r_1}\,d\theta_1,\end{aligned} \tag{7}$$

and it is clear that all terms in the integrand have a finite limit as the field point approaches the source position.

The singular kernel must be treated differently to extract the leading order singularity. The aim is to make use of existing results for axisymmetric systems, as robust techniques are available for solving such problems. Expanding $g_m^{(s)}$ as a power series

$$\frac{\partial^2 g_m^{(s)}}{\partial r\,\partial r_1} = \sum_{u=0}^{\infty}\frac{\epsilon^u}{u!}\frac{\partial^u}{\partial\epsilon^u}\left(\frac{\partial^2 g_m^{(s)}}{\partial r\,\partial r_1}\right)\Bigg|_{\epsilon=0} \tag{8}$$

which can be rewritten

$$\frac{\partial^2 g_m^{(s)}}{\partial r\,\partial r_1} = \sum_{u=0}^{\infty}\frac{(\epsilon\Delta w)^u}{u!}\frac{\partial^u}{\partial z^u}\left(\frac{\partial^2 g_m^{(s)}}{\partial r\,\partial r_1}\right)\Bigg|_{\epsilon=0}, \tag{9}$$

because z and ϵ only occur in the combination $(z-z_1+\epsilon\Delta w)$.

Because $\epsilon=0$ corresponds to the axisymmetric problem, the singular kernel can be computed using established results. The terms which must be handled are of the form $\exp jm\theta_1/R_\epsilon^n$ which can be written

$$\frac{1}{R_\epsilon^2} = \sum_{u=0}^{\infty} \frac{(\epsilon \Delta w)^u}{u!} \frac{\partial^\mu}{\partial z^u} \frac{1}{R_0^n}, \tag{10}$$

where R_0 is R_ϵ evaluated at $\epsilon=0$. The terms related to the end profile $w(\theta)$ are expanded in a Fourier series:

$$(\Delta w)^u = \sum_{p,q} W_{pq}^{(u)} e^{jp\theta} e^{jp\theta_1},$$

where nonzero values of $W_{pq}^{(u)}$ are included in the summation and

$$\frac{1}{R_\epsilon^n} = \sum_{u=0}^{\infty} \frac{\epsilon^u}{u!} \frac{\partial^\mu}{\partial z^u} \sum_{p,q} W_{pq}^{(u)} e^{jp\theta} \frac{e^{jq\theta_1}}{R_0^n}. \tag{11}$$

Integration of Eq. (11) yields

$$\int_0^{2\pi} \frac{e^{jm\theta_1}}{R_\epsilon^n} d\theta_1 = \sum_{u=0}^{\infty} \frac{\epsilon^u}{u!} \frac{\partial^\mu}{\partial z^u} \sum_{p,q} W_{pq}^{(u)} e^{jp\theta} \int_0^{2\pi} \frac{e^{j(m+q)\theta_1}}{R_0^n} d\theta_1.$$

This can be simplified as follows. Using the substitution $\theta_1' = \theta - \theta_1$

$$\int_0^{2\pi} \frac{e^{jm\theta_1}}{[r^2 + r_1^2 - 2rr_1 \cos(\theta - \theta_1) + (z - z_1)^2]^{n/2}} d\theta_1$$

$$= e^{jm\theta} \int_0^{2\pi} \frac{\cos m\theta_1'}{[r^2 + r_1^2 - 2rr_1 \cos\theta_1' + (z - z_1)^2]^{n/2}} d\theta_1'$$

$$= e^{jm\theta} \int_0^{2\pi} \frac{\cos 2m\phi}{[(r + r_1)^2 + (z - z_1)^2 - 4rr_1 \cos^2\phi]^{n/2}} d\phi$$

after the substitution $\theta' = 2\phi$. Then, setting

$$\rho^2 = (r + r_1)^2 + (z - z_1)^2, \quad \lambda^2 = \frac{4rr_1}{\rho^2},$$

and defining

$$H(n,m,\lambda) = \int_0^{2\pi} \frac{\cos 2m\phi}{(1 - \lambda^2 \cos^2\phi)^{n/2}} d\phi, \tag{12}$$

Eq. (11) becomes

$$\int_0^{2\pi} \frac{e^{jm\theta_1}}{R_\epsilon^n} d\theta_1 = \sum_{u=0}^{\infty} \frac{\epsilon^u}{u!} \frac{\partial^\mu}{\partial z^u} \sum_{p,q} W_{pq}^{(u)} e^{j(p+q+m)\theta} \frac{H(n,m+q,\lambda)}{\rho^n}. \tag{13}$$

The elliptic-type integral $H(n,m,\lambda)$ has been extensively analyzed due to its importance in radiation problems. In particular, recursion relations have been derived for its efficient evaluation and an asymptotic series exists for the case where $n=2n'+1$, covering the cases of interest in this paper:[8]

$$H(2n'+1,m,\lambda) = \frac{\lambda^{2m} \lambda'^{-2n'}}{(1/2)_{n'}} \times \left(\sum_{k=0}^{n'-1} (-1)^k C_k (n' - 1 - k)!\, \lambda'^{2k} + (-1)^{n'} \sum_{k=n'}^{\infty} \frac{C_k}{(k - n')!} h_k(\lambda') \lambda'^{2k} \right), \tag{14}$$

where

$$C_k = \left(\frac{1}{2} + m\right)_k \left(\frac{1}{2} + m - n'\right)_k / k!,$$

$$h_k(\lambda') = \psi(1 + k) + \psi(1 + k - n') - \psi\left(\frac{1}{2} + m + k\right) - \psi\left(\frac{1}{2} + m - n' + k\right) - 2 \log \lambda',$$

$$\lambda' = (1 - \lambda^2)^{1/2},$$

ψ is the logarithmic derivative of the gamma function, and $(a)_n$ is Pochammer's symbol. The asymptotic series, as well as being useful in evaluating the kernel proper, is also useful in that it gives the strength of the leading order singularity as $\lambda \to 1$.

Differentiating Eq. (6c) with $\epsilon=0$,

$$K_m^{(s)} = \left. \frac{\partial^2 g_m^{(s)}}{\partial r\, \partial r_1} \right|_{\epsilon=0} = \frac{1}{8\pi} \frac{r^2 + r_1^2}{rr_1} \frac{H(3,m,\lambda)}{\rho^3} + \frac{3}{16\pi} \frac{1}{rr_1} [(z - z_1)^4 - (r - r_1)^2 (r + r_1)^2] \times \frac{H(5,m,\lambda)}{\rho^5} + \frac{1}{8\pi} \frac{1}{rr_1} [r^2 + r_1^2 - 2(z - z_1)^2] \times \frac{H(3,m,\lambda)}{\rho^3} + \frac{1}{16\pi} \frac{1}{rr_1} \frac{H(1,m,\lambda)}{\rho^1} + \frac{k^2}{32\pi} \frac{1}{rr_1} [(z - z_1)^4 - (r - r_1)^2 (r + r_1)^2] \frac{H(3,m,\lambda)}{\rho^3} + \frac{k^2}{16\pi} \frac{1}{rr_1} (r + r_1)^2 \frac{H(1,m,\lambda)}{\rho^1} - \frac{k^2}{32\pi} \frac{1}{rr_1} \frac{H(-1,m,\lambda)}{\rho^{-1}} - \frac{k^2}{8\pi} \frac{1}{rr_1} \frac{H(1,m,\lambda)}{\rho^1}, \tag{15}$$

so that the asymmetric kernel becomes

$$\frac{\partial^2 g_m^{(s)}}{\partial r\, \partial r_1} = \sum_{u=0}^{\infty} \frac{\epsilon^u}{u!} \frac{\partial^\mu}{\partial z^u} \sum_{p,q} W_{pq}^{(u)} e^{j(p+q+m)\theta} K_{m+q}^{(s)}. \tag{16}$$

We note in passing that the substitution $r=r_1 \equiv 1$ has not been made in Eq. (15) in order to allow the numerical procedures to be checked by evaluating the kernel with $r \neq r_1$.

Finally, inserting Eq. (16) into the integral equation and moving the differentiations from under the integral sign

$$\int_{-1}^{1} \frac{\partial^2 g_m^{(s)}}{\partial r\, \partial r_1} U_n(t)(1-t^2)^{1/2}\, \mathrm{d}t$$

$$= \sum_{u=0}^{\infty} \frac{\epsilon^u}{u!} \frac{\partial^u}{\partial z^u} \sum_{p,q} W_{pq}^{(u)} \mathrm{e}^{\mathrm{j}(p+q+m)\theta} \times \int_{-1}^{1} K_{m+q}^{(s)} U_n(t)(1-t^2)^{1/2}\, \mathrm{d}t. \tag{17}$$

C. Hypersingular kernel

The singular part of the kernel requires special treatment as standard numerical integration techniques will not work. The first step is to isolate the most singular term. This can be done most easily using the asymptotic series, Eq. (14). With $r=r_1 \equiv 1$, the most strongly singular term is the first term in

$$\frac{1}{4\pi} \frac{H(3,m,\lambda)}{\rho^3}.$$

Setting

$$\lambda'^2 = \frac{4}{4+(z-z_1)^2},$$

the leading order singularity is

$$\frac{1}{4\pi} \frac{H(3,m,\lambda)}{\rho^3} \sim \frac{1}{2\pi} \frac{1}{(z-z_1)^2}, \quad z \to z_1,$$

which can also be derived using small argument methods.[6]

This result means that the integrand has hypersingular behavior as $z \to z_1$ and the integral of Eq. (5) must be interpreted in the Hadamard finite-part sense.[14–16] A number of approaches exist for the evaluation of finite-part integrals but the most convenient for our purposes is the Gaussian-type quadrature of Korsunsky,[7] which is a special case of the rules derived by Monegato.[17] This quadrature is particularly good for integral equations based on Chebyshev polynomial expansions where the kernel is difficult to analyze in detail. When the kernel can be written in the form

$$K(t,x) = \frac{1}{(t-x)^2} + k(t,x),$$

the quadrature rule of order n is

$$\fint_{-1}^{1} K(t,x) g(t)(1-t^2)^{1/2} \mathrm{d}t \approx \pi \sum_{i=1}^{n} \left(\frac{1-t_i^2}{n+1} K(t_i,x_k) + \frac{1-t_i^2}{t_i-x_k} \frac{(-1)^{i+k}}{(1-x_k^2)^{1/2}} \right) g(t_i), \tag{18}$$

$$t_i = \cos\frac{i\pi}{n+1} \quad \text{and} \quad x_k = \cos\frac{(2k-1)\pi}{2(n+1)}, \quad k = 1, \ldots, n+1,$$

where $\fint$ denotes the finite-part integral. Note that $k(t,x)$ may contain singularities up to order 1 and these will be evaluated as Cauchy principal values.

D. Differentiation of kernel

Equation (17) requires that the integral in it be differentiated with respect to the axial coordinate. The collocation points are defined by Eq. (18) and lie in the range $-1 < x < 1$ where x is the transformed variable inside the range of integration. Differentiation with respect to the physical variable z is then simply a scaled version of differentiation with respect to x:

$$\frac{\partial}{\partial z} = \frac{2}{L} \frac{\partial}{\partial x}.$$

In assembling the equations, I is evaluated at the collocation points so that we start from a set of data (x_k, I_k). In order to perform the differentiation, we fit a set of Chebyshev polynomials of the first kind to the data and use the coefficients of the series to evaluate the derivative. The Chebyshev expansion of a function is given by

$$I(x) = \sum_{m=0}^{m=M} a_m T_m(x), \tag{19}$$

where, by the orthogonality relation for Chebyshev polynomials,[18]

$$a_0 = \frac{1}{\pi} \int_{-1}^{1} I(x)(1-x^2)^{-1/2}\, \mathrm{d}x,$$

$$a_m = \frac{2}{\pi} \int_{-1}^{1} I(x) T_m(x)(1-x^2)^{-1/2}\, \mathrm{d}x, \quad m \neq 0.$$

Because of the location of the collocation points x_k, it is convenient to transform the integrals for the Chebyshev coefficients. Setting $x = \cos\phi$ and noting that $T_m(x) \equiv \cos m\phi$,

$$a_m = \frac{2}{\pi} \int_0^{\pi} I(\cos\phi) \cos m\phi\, \mathrm{d}\phi.$$

The integral has been evaluated at the collocation points given by Eq. (18) so that the data are equally spaced in ϕ and

$$a_0 \approx \frac{1}{2n} \sum_{k=0}^{n} I_k, \tag{20a}$$

$$a_m \approx \frac{1}{n} \sum_{k=0}^{n} I_k T_m(x_k), \quad m \neq 0. \tag{20b}$$

To evaluate the derivative, we use the formula $\mathrm{d}T_m/\mathrm{d}x = mU_{m-1}(x)$ and

$$\frac{\partial I}{\partial x} = \sum_{m=0}^{M} m a_m U_{m-1}(x), \tag{21}$$

which can be differentiated in turn to evaluate the higher derivatives.

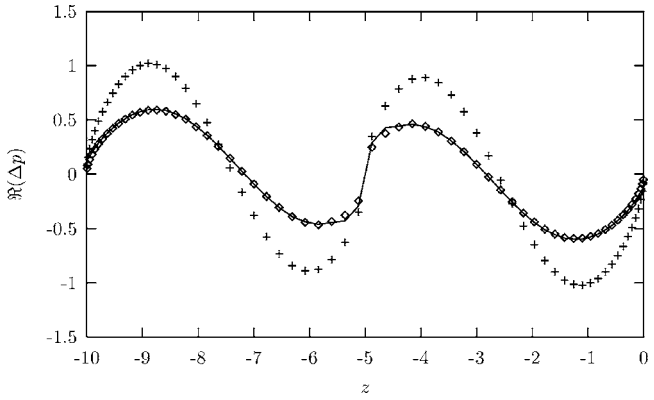


FIG. 2. Convergence of solution for scarf angle 30° (ϵ=0.58): surface pressure jump $\mathcal{R}(\Delta p)$ against the axial coordinate at $\theta=\pi/2$; cross: first order; diamond: third order; solid line: fifth order.

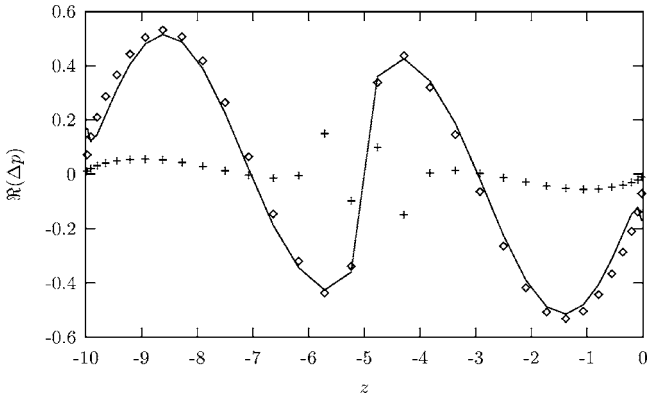


FIG. 3. Convergence of solution for scarf angle 45° (ϵ=1.0): surface pressure jump $\mathcal{R}(\Delta p)$ against the axial coordinate at $\theta=\pi/2$; cross: first order; diamond: third order; solid line: fifth order.

E. Implementation

The implementation of the method follows the sequence given above. The incident field azimuthal order is specified as are the azimuthal orders of the modes to be included in the scattering calculation. For a field of order m incident on a linear scarf, the required modes will be $m\pm u$, where $u=0,1,\ldots,U$ where, in the results shown here, U is the maximum order of the expansion in ϵ. The profile $w(\theta)$ is specified, allowing $W_{pq}^{(u)}$ to be computed using a two-dimensional Fourier transform.[19]

For M azimuthal modes, the boundary conditions are imposed at angular positions $\theta_i=i2\pi/M$, where $i=0,\ldots,M-1$, and at axial positions found from the quadrature rule, Eq. (18):

$$z_k=\bar{z}+\frac{L}{2}x_k+\epsilon w(\theta_i).$$

This yields a well-posed system of equations for Q_{nm} which can be used in Eq. (1) to calculate the scattered pressure.

III. RESULTS

Before presenting some sample results of the method, the numerical performance of the technique is analyzed. The important property of the method is its convergence behavior with respect to ϵ and the order of the solution, i.e., the highest power of ϵ considered. Figure 2 presents results for a linear scarf with ϵ=0.58 and k=1.0, computed using terms up to $O(\epsilon)$, $O(\epsilon^3)$, and $O(\epsilon^5)$, a case considered in Sec. III A. The number of azimuthal modes included in the solution is increased as the order of the calculation rises. Figure 2 shows the real part of the computed surface pressure jump $\mathcal{R}(\Delta p)$ against the surface axial coordinate. The solid curve shows the $O(\epsilon^5)$ solution while the diamonds and crosses show the $O(\epsilon^3)$ and $O(\epsilon)$ solutions, respectively. It can be seen that the third- and fifth-order solutions coincide. Figure 3 shows equivalent data for ϵ=1, the most extreme case computed in this paper, and the behavior with ϵ is seen to be similar: poor results if only terms up to $O(\epsilon)$ are included but convergence thereafter. The behavior of the solution for smaller ϵ was similar and so all of the results which follow have been computed to order ϵ^5.

The second issue is that of the time required to solve a problem. Figure 4 shows the computational time required to assemble the matrix for solution of the linear scarf problem at ϵ=1 with different numbers of modes included. The data were generated by increasing the order of ϵ and the number of modes simultaneously. The computational time shown has been scaled on the time required in the single mode case, i.e., the axisymmetric problem. There is no clear-cut variation in computational time with the number of modes N, but it does appear to be roughly N^2, as shown by the superimposed quadratic fit.

A. Plane wave-linear scarf interaction

The first problem considered is that of radiation from a pipe with a linear scarf. In recent papers, asymptotic analysis has been applied to low-frequency scattering by ducts with a small scarf angle[1] and ray theory to high-frequency radiation from ducts of arbitrary scarf angle.[10] It has been noted,[1] however, that a numerical solution is required when both the scarf angle and the wave number are $O(1)$. For this reason, the nondimensional wave number k=1 and the radiated field is calculated for scarf angles of 0°, 15°, 30°, and 45° ($\epsilon=0,0.27,0.58,1.0$). The incident field is generated by a ring of axial dipoles of unit strength at radius $\frac{1}{2}$ placed at $z=-5$ in a duct of length 10 with $a=-10$ and $b=0$. The next cut-on

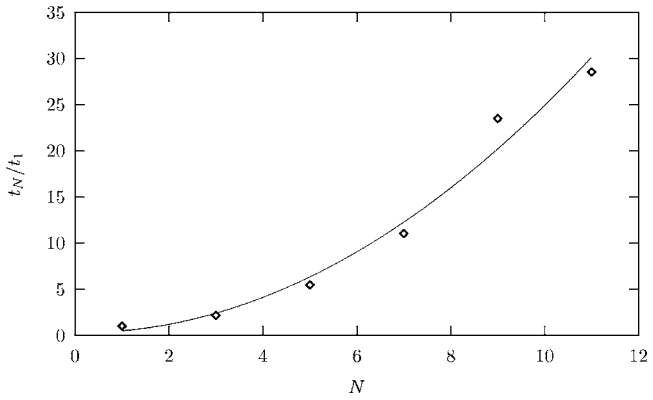


FIG. 4. Computational time for matrix assembly against number of modes included. Time has been scaled on time for single mode (axisymmetric problem). Diamonds: computation time; solid line: quadratic polynomial fit.

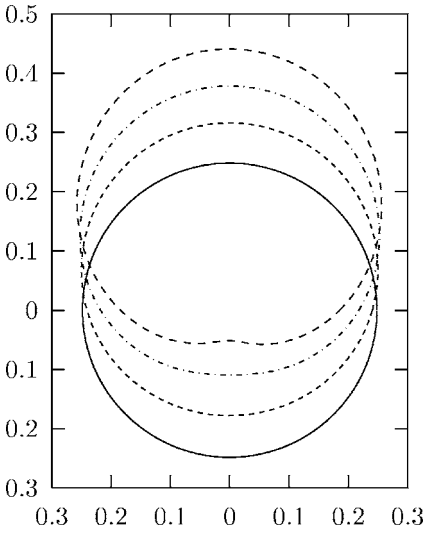


FIG. 5. Polar plot of $|p|^2R^2$ against θ at $\phi=\pi/4$ for an axisymmetric field incident on a linear scarf. Wave number $k=1.0$; scarf angles 0°, 15°, 30°, and 45°, shown as solid, dashed, dot-dashed, and dashed lines, respectively.

wave number is $k=3.832$ so that only the plane mode propagates.[20] Far-field directivities are plotted as $|p|^2R^2$ where the radial coordinate R is centered on the nominal center of the duct face at $z=0$. In performing the calculation $R=20$.

Figure 5 shows the normalized sound power at constant $\phi=\pi/4$ with azimuthal angle θ varying. The solid line represents the axisymmetric case $\epsilon=0$ and the directivity pattern moves up as the scarf angle increases. As in Fig. 3 of Peake's work,[1] the scarfing has shifted the directivity so that the noise below the duct is reduced and that above it increased—precisely the aim of scarfing.

Figure 6 shows the directivity in a vertical plane through the duct centerline. The field ahead of the inlet plane is composed of a narrow lobe near the vertical and a broader one along the axis. Behind the inlet plane, there is a second set of lobes. These are the sound field radiated from the rear of the pipe (which is negatively scarfed) as can be seen by computing the directivity angle from the inlet to the rear plane which is 120°. As expected from previous work, increased scarf angle reduces the amplitude of the main lobe below the

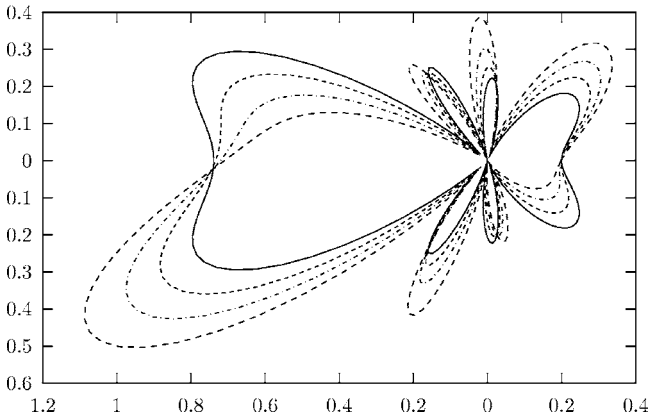


FIG. 6. Polar plot of $|p|^2R^2$ against ϕ at $\theta=\pi/4$ for an axisymmetric field incident on a linear scarf. Wave number $k=1.0$; scarf angles 0°, 15°, 30°, and 45°, shown as solid, dashed, dot-dashed, and dashed lines, respectively.

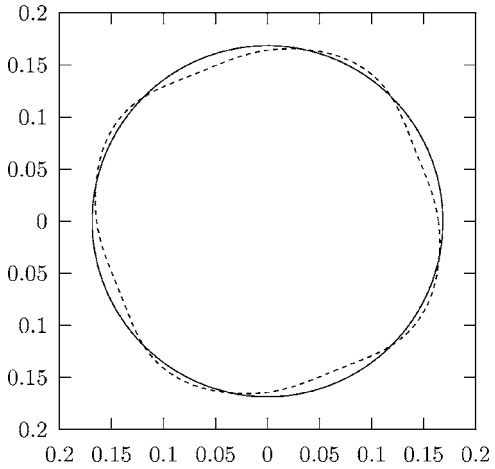


FIG. 7. Polar plot of $|p|^2R^2$ against θ at $\phi=\pi/4$ for field of azimuthal order $m=4$ incident on a four-lobed termination, $k=5$, $\epsilon=0$, 0.25, shown as solid and dashed, respectively.

pipe axis at the expense of an increase above it. Likewise, the narrower lobe in the vertical plane increases above the horizontal but does not reduce significantly below, merely being shifted slightly. The conclusions drawn in earlier work on scarfing are thus confirmed for the large scarf angle considered here.

B. Azimuthal mode-lobed nozzle interaction

The second set of results is a calculation of the sound radiated by a mode of nonzero azimuthal order incident on a sinusoidally profiled duct termination, a simple model of a chevron mixer.[2] In this case, $k=5$, which roughly corresponds to sound generated at a frequency of 550 Hz near a nozzle of diameter 1 m. The nozzle profile is $\sin 4\theta$ and its amplitude is $\epsilon=0.25$. As a baseline condition, the source is an axial dipole distribution of radius 1 with azimuthal variation of order $m=4$, positioned outside the nozzle at $z=1$, i.e., one nozzle radius "downstream." For comparison, the same calculation is also performed for modes $m=3$ and $m=5$. As in Fig. 5, the results are plotted as polar directivities $|p|^2R^2$ at $\phi=\pi/4$, with the axisymmetric case superimposed for comparison.

Figures 7–9 show the baseline case $m=4$ and the comparisons at $m=3$ and $m=5$, respectively. The first point is that the lobing, although it increases the radiated sound slightly, has little effect in the $m=4$ and $m=5$ cases. In the $m=3$ case, however, the increase is quite large and there is practically no reduction in any direction.

The reason for this appears to be that the $m=3$ mode is cut-on in the duct at $k=5$, while the $m=4$ and $m=5$ modes are cut-off. The propagatating mode generates a strong scattering response, including the cut-on $m=-1$ mode, which contributes the asymmetry which is apparent in Fig. 8. By contrast, in the other two cases, although the incident modes excite cut-on modes, they do so via a mode which is cut-off in the duct so that the amplitude of the excited modes is quite weak and no great asymmetry is apparent.

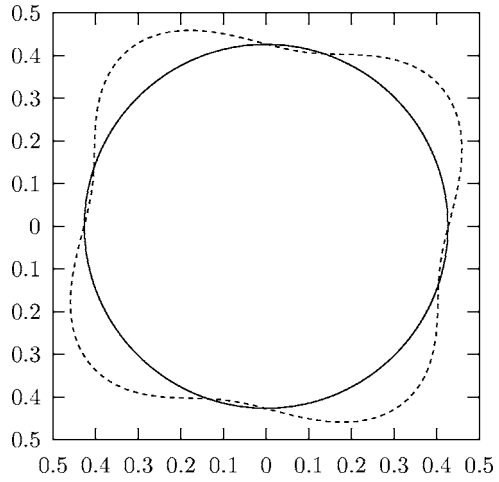


FIG. 8. Polar plot of $|p|^2R^2$ against θ at $\phi=\pi/4$ for field of azimuthal order $m=3$ incident on a four-lobed termination, $k=5$, $\epsilon=0$, 0.25, shown as solid and dashed, respectively.

IV. CONCLUSIONS

A method has been presented for the prediction of scattering by straight circular pipes with arbitrary end profiles. The method has been applied to the problem of radiation of sound from a linearly scarfed duct and to that of scattering by a profile representative of a chevron nozzle on an aeroengine. In the case of the scarfed duct, it was found that there is a limit on the noise reduction benefit to be had by scarfing beyond a certain point. Examination of the sample results for a chevron nozzle has shown that the purely acoustic effect of the end profile depends on the interaction between the internal modes of the pipe and the incident field. The procedure has proven simple to implement using recently developed numerical methods. Future developments might include using a similar approach to model scattering by pipes of almost circular section, or to predict radiation from ducts with asymmetric liners.

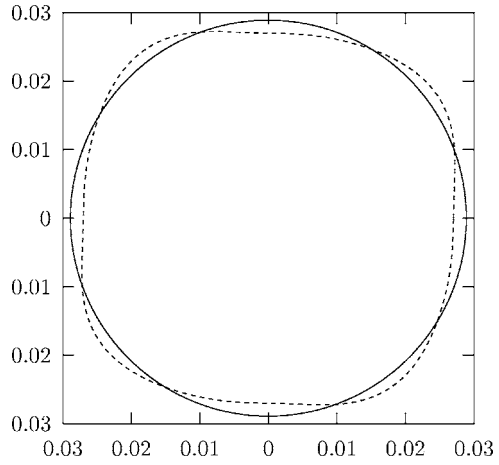


FIG. 9. Polar plot of $|p|^2R^2$ against θ at $\phi=\pi/4$ for field of azimuthal order $m=5$ incident on a four-lobed termination, $k=5$, $\epsilon=0$, 0.25, shown as solid and dashed, respectively.

[1]N. Peake, "On the radiation properties of an asymmetric cylinder," Wave Motion **22**, 371–385 (1995).
[2]J. Bridges and C. A. Brown, "Parametric testing of chevrons on single flow hot jets," in Tenth AIAA/CEAS Aeroacoustics Conference (AIAA, Manchester, UK, 2004), No. AIAA-2004-2824.
[3]R. M. Munt, "The interaction of sound with a subsonic jet issuing from a semi-infinite cylindrical pipe," J. Fluid Mech. **83**(4), 609–640 (1997).
[4]R. Martinez, "Diffracting open-ended pipe treated as a lifting surface," AIAA J. **26**(4), 396–404 (1988).
[5]R. Martinez, "A boundary integral formulation for thin-walled shapes of revolution," J. Acoust. Soc. Am. **87**(2), 523–531 (1990).
[6]M. H. Dunn, J. Tweed, and F. Farassat, "The application of a boundary integral equation method to the prediction of ducted fan engine noise," J. Sound Vib. **227**(5), 1019–1048 (1999).
[7]A. M. Korsunsky, "Gauss-Chebyshev quadrature formulae for strongly singular integrals," Q. Appl. Math. **56**(3), 461–472 (1998).
[8]J. Björkberg and G. Kristensson, "Electromagnetic scattering by a perfectly conducting elliptic disk," Can. J. Phys. **65**, 723–734 (1987).
[9]J. M. Tyler and T. G. Sofrin, "Axial flow compressor noise studies," Trans. Soc. Automotive Eng. **70**, 309–332 (1962).
[10]G. M. Keith and N. Peake, "High-wavenumber acoustic radiation from a thin-walled scarfed cylinder," J. Sound Vib. **255**(1), 147–160 (2002).
[11]D. B. Hanson, "Sound from a propeller at angle of attack: a new theoretical viewpoint," Proc. R. Soc. London, Ser. A **449**, 315–328 (1995).
[12]M. Carley, "The structure of wobbling sound fields," J. Sound Vib. **244**(1), 1–19 (2001).
[13]T. W. Dawson, "On the singularity of the axially symmetric Helmholtz Green's function, with application to BEM," Appl. Math. Model. **19**, 590–599 (1995).
[14]K. W. Mangler, "Improper integrals in theoretical aerodynamics," Current Paper 94, Aeronautical Research Council (unpublished).
[15]M. J. Lighthill, *An Introduction to Fourier Analysis and Generalised Functions* (Cambridge University Press, Cambridge, 1958).
[16]M. P. Brandão, "Improper integrals in theoretical aerodynamics: The problem revisited," AIAA J. **25**(9), 1258–1260 (1987).
[17]G. Monegato, "Numerical evaluation of hypersingular integrals," J. Comput. Appl. Math. **50**, 9–31 (1994).
[18]I. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 5th ed. (Academic, London, 1980).
[19]M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," Proc. IEEE **93**(2), 216–231 (2005).
[20]A. D. Pierce, *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York, 1989).

Wind-induced noise in a screened microphone

G. P. van den Berg[a)]
Science Shop for Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

(Received 21 April 2005; revised 4 November 2005; accepted 7 November 2005)

To reduce the effect of wind on a microphone, wind screens are used, usually made of foam or cloth. Although this measure to reduce wind noise is well known and widely used, it lacks a theoretical explanation. One possible explanation was the turbulent wake in the air flow behind the wind screen, another the pressure variations due to turbulence inherent in outdoor wind. In the present paper it is shown that atmospheric turbulence is indeed the cause of outdoor microphone wind noise, and the knowledge that is available in atmospheric boundary layer physics can be used to model the effects of atmospheric turbulence on a bare or screened microphone. A wind screen can be seen as a first order low pass filter for pressure variations due to atmospheric turbulence. Induced wind noise is thus determined by (average) wind speed and wind screen diameter, but also by factors related to frictional and thermal turbulence (roughness length and atmospheric stability). This approach leads to predicted spectral values of the induced "sound" pressure that match experimental data. The result can be used to obtain proper outdoor ambient wind related sound levels by correcting measured sound levels with calculated wind-induced pressure levels. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146085]

PACS number(s): 43.28.Ra, 43.28.Vd, 43.38.Kb, 43.50.Rq [GCL] Pages: 824–833

I. INTRODUCTION

It is commonly known that a wind screen over a microphone reduces "wind noise" that apparently results from the air flow around the microphone. An explanation for this phenomenon has been addressed by several authors. According to a dimensional analysis by Strasberg[1] the pressure within a spherical or cylindrical wind screen with diameter D in a flow with velocity V depends on Strouhal number $\mathrm{Sr}=fD/V$, Reynolds number $\mathrm{Re}=DV/\nu$, and Mach number $\mathrm{M}=V/c$ (where ν is the kinematic viscosity of air and c the velocity of sound). Writing the rms pressure in a relatively narrow frequency band centered at frequency f as p_f, and in dimensionless form by division with ρV^2 (ρ is air density), Strasberg found $p_f/\rho V^2=\mathrm{function}(\mathrm{Sr},\mathrm{Re},\mathrm{M})$. Comparison with measured 1/3 octave band levels from four authors on 2.5–25 cm diameter wind screens, in air speeds ranging from 6 to 23 m/s yielded a definite expression for 1/3 octave frequency band pressure levels re ρV^2

$$L_{p,1/3}=20\log_{10}(p_{1/3}/\rho V^2)=-23\log_{10}(f_m D/V)-81, \quad (1)$$

where f_m is the middle frequency of the 1/3 octave band. The data points agreed within approximately 3 dB with Eq. (1) for $0.1<fD/V<5$, except for one of the 14 data series where measured values diverged at $fD/V>2$. Equation (1) can also be written in acoustical terms by expressing the rms pressure as a sound pressure level relative to 20 μPa

$$L_{1/3}=40\log_{10}(V/V_0)-23\log_{10}(f_m D/V)+15. \quad (2)$$

Here V_0 is a reference velocity of 1 m/s and $\rho=1.23\ \mathrm{kg/m^3}$ is used (air density at 1 bar and 10 °C). Equation (2) is slightly different from the expression given by Strasberg because SI units are used and terms in logarithms have been nondimensionalized.

Morgan and Raspet pointed out that all measurements reported by Strasberg were made in low turbulence flows, such as wind tunnel flow.[2] Strasberg's result thus referred to the wake created by a wind screen and excluded atmospheric turbulence (as Strasberg had noted himself in his concluding remarks[1]). Outdoors, however, the flow is turbulent, and induced pressure variations are expected to depend on meteorological parameters also. Morgan and Raspet applied Bernoulli's principle by decomposing the wind velocity U in a constant time-averaged velocity V and a fluctuation velocity u with a time average $\langle u\rangle=0$, to obtain the rms pressure fluctuation $\mathbf{p}=\rho V\mathbf{u}$ (a bold character is used to denote the rms value $\mathbf{x}$ of a variable x: $\mathbf{x}=\sqrt{\langle x^2\rangle}$). This method can be compared to Strasberg's model for a microphone in turbulent water flow.[16] Measurements in wind speeds of 3–13 m/s at 30.5 and 1.5 m height for different screen diameters (90 and 180 mm) and screen pore sizes (10, 20, 40, and 80 ppi) yielded

$$\mathbf{p}=\alpha\cdot\rho(V\mathbf{u})^k, \quad (3)$$

with α ranging from 0.16 to 0.26 and k from 1.0 to 1.3.[2] For some measurements Morgan and Raspet showed spectra over almost the same frequency range where Eq. (1) is valid ($0.1<fD/V<5$). At frequencies higher than 3 Hz the slope is roughly comparable to what Strasberg found, but values of $L_{p,1/3}$ are 8–20 dB higher then expected by Eq. (1), implying that atmospheric turbulence dominated expected wake turbulence.

Zheng and Tan tried to solve this problem analytically.[3] Their analysis applies to low frequency variations, when the velocity variation u is uniform over the wind screen. Zheng and Tan state that this assumption seems to be valid for a low screen number $D/\lambda(<0.3)$, the ratio between screen diam-

a)Electronic address: g.p.van.den.berg@rug.nl

eter and wavelength. Ignoring viscous effects (i.e., Reynolds number is infinite), and calculating the pressure variation at the center of a spherical wind screen caused by pressure variations at the surface induced by a wind speed $U=V+u$, they found $p=-\frac{1}{2}\cdot\rho Vu$ or

$$\mathbf{p}=\tfrac{1}{2}\rho V\mathbf{u}. \tag{4}$$

Comparison with Eq. (3) shows that now $\alpha=0.5$ and $k=1$. For finite Reynolds numbers ($\mathrm{Re}/10^4\approx 0.5-15$ for wind screens of 4–20 cm and wind speeds of 2–12 m/s), screening is better,[3] and $\alpha\leqslant 0.5$.

Zheng and Tan argued that the effect of turbulent velocity fluctuations perpendicular to the average wind speed, in the vertical (w) as well as horizontal (v) direction, can be neglected in a first-order approximation, as it scales with v^2 and w^2 and is therefore second order compared to the effect of the component u in line with the average wind speed V that scales as Vu.[3]

Finally Boersma showed the agreement between sound pressure spectra due to wind measured at 1.5 m above flat, open grassland and Strasberg's results.[4] However, Boersma used 95 percentile levels (L_{95}) which he estimated to be 6 to 13 dB lower than equivalent sound levels in the range considered ($30<L_{95}<70$ dB),[4] but he did not apply a level correction. So in fact he found that his wind related spectra had slopes comparable to Strasberg's, but with a 6–13 dB higher value, not unlike the Morgan and Raspet spectra.

In short, from the literature it is clear that air turbulence creates pressure fluctuations especially at low frequencies, but the origin—wake or atmospheric turbulence—has not been definitely resolved.

In this paper a quantitative model will be presented for the level of pressure variations due to atmospheric turbulence, i.e., the "sound" pressure level taken from a sound level meter caused by turbulence on the microphone wind screen. First the spectral distribution of atmospheric turbulence and its effect on a screened microphone will be described (Sec. II). These levels will be compared to measured spectra related to wind, obtained by the author as well as by others (Sec. III). Finally the results will be discussed (Sec. IV).

II. ATMOSPHERIC TURBULENCE

Writing the velocity U as the sum of a constant (average) wind speed V and a fluctuating part u, and similarly $P=P_{\mathrm{average}}+p$, the relation between the rms microphone pressure fluctuation $\mathbf{p}$ and the rms wind velocity fluctuation $\mathbf{u}$ is $\mathbf{p}=\alpha\rho V\mathbf{u}$. The pressure level due to atmospheric turbulence can be expressed as a sound pressure level L_{at} (with reference pressure $p_{\mathrm{ref}}=20$ μPa)

$$L_{\mathrm{at}}(\mathbf{u})=20\log_{10}\left(\frac{\alpha\rho V\mathbf{u}}{p_{\mathrm{ref}}}\right), \tag{5}$$

which is frequency dependent because of $\mathbf{u}$.

A. Turbulence spectra

Atmospheric turbulence is treated in many papers and textbooks (such as Jensen and Busch[5] or Zhang *et al.*[6]), also in reference to acoustics (see, e.g., Wilson and Thomson[7]). Here a short elucidation will be presented, leading to our topic of interest: turbulence spectra.

Atmospheric turbulence is created by friction and by thermal convection. Turbulence due to friction is a result of wind shear: At the surface the wind speed is zero whereas at high altitudes the geostrophic wind is not influenced by the surface but a result of large scale pressure differences as well as Coriolis forces resulting from the Earth's rotation. In between, in the atmospheric boundary layer, wind speed increases with height z (see Ref. 8, p. 53)

$$V=(u*/\kappa)[\ln(z/z_0)-\Psi]. \tag{6}$$

Here $\kappa=0.4$ is von Karman's constant, z_0 is the roughness height and u_* is the friction velocity, defined by $u_*^2=\sqrt{[\langle uw\rangle^2+\langle uw\rangle^2]}=\tau/\rho$, where τ equals the momentum flux due to turbulent friction across a horizontal plane.

$\Psi=\Psi(\zeta)$ is a function (of $\zeta=z/L$) correcting for atmospheric stability. Monin Obukhov length L is an important length scale for stability and can be thought of as the height above which thermal turbulence dominates over friction turbulence. Garratt gives the following approximations:[8]

- In a stable atmosphere $L>0$ and $\Psi(\zeta)=-5\zeta<0$.
- In a neutral atmosphere $|L|$ is large or $1/L\approx 0$, $\Psi(0)\approx 0$, and Eq. (6) reduces to the well-known logarithmic wind profile.
- In an unstable atmosphere $L<0$ and $\Psi(\zeta)=2\ln[(1+x)/2]+\ln[(1+x^2)/2]-2/\tan(x)+\pi/2>0$, where $x=(1-16\cdot\zeta)^{1/4}$.

For $-1<\zeta<1$, $\Psi(\zeta)$ is of the same order of magnitude as the logarithmic term in Eq. (6) for common values of measurement height ($1<z<5$ m) and roughness length ($1<z_0<10$ cm).

The friction created by wind shear produces eddies over a range of frequencies and lengths, their size determined by z and V. These eddies break up in ever smaller eddies and kinetic turbulent energy is cascaded to smaller sizes at higher frequencies until the eddies reach the Kolmogorov size η_s (≈ 1 mm) and dissipate into heat by viscous friction. It has been shown by Kolmogorov that for this energy cascade, in the so-called inertial subrange of the turbulent spectrum, the frequency dependency follows the "law of 5/3:" the spectrum falls with $f^{-5/3}$. It is customary in atmospheric boundary layer physics to express turbulence frequency in dimensionless form n, with $n=fz/V$ (in fact n and f are usually interchanged, but in this paper f will be used for dimensional frequency, as is usual in acoustics). The seminal Kansas measurements[9] showed that the rms longitudinal velocity fluctuation $\mathbf{u}_f$ per unit frequency depends on frequency as

$$\frac{f\cdot\mathbf{u}_f^2}{u_*^2}=105n(1+33n)^{-5/3}. \tag{7}$$

The experimentally determined constants in this equation, the nondimensional turbulent energy spectrum, are not exact, but are close to values determined by others.[6,8] For $n\ll 1/33$, the right-hand side approximates $105n$, which, with $n=fz/V$ and Eq. (6), leads to $\mathbf{u}_f^2=105\cdot u_*^2\cdot z/V$

$=105\kappa^2 zV[\ln(z/z_0)-\Psi]^{-2}$. Applying this to Eq. (5), the induced pressure level per unit of frequency appears to be independent of frequency, but increases with wind speed ($\sim 30 \log V$). This is the turbulent energy producing subrange.

For $n \gg 1/33$ the right-hand side of Eq. (7) reduces to $3.2(33n)^{-2/3}$, leading to $\mathbf{u}_f^2 = 0.3 \cdot u_*^2 (V/z)^{2/3} f^{-5/3}$, which describes the inertial subrange. The frequency where Eq. (7) has a maximum is $n_{\max}=0.05$ or $f_{\max}=0.05\ V/z$. As sound measurement are usually at heights $1 \leqslant z \leqslant 5$ m, $f_{\max}$ is less than 1 Hz for wind speeds $V<20$ m/s and thus below the acoustically relevant frequencies.

When insolation increases the surface temperature, the atmosphere changes from neutral to unstable and eddies are created by thermal differences with sizes up to the boundary layer height with an order of magnitude of 1 km. Turbulent kinetic energy production then shifts to lower frequencies. In contrast to this, in a stable atmosphere, where surface temperature decreases because of surface cooling, eddy production at low frequencies (corresponding to large eddy diameters) is damped and the spectral maximum shifts to a higher frequency up to $n \approx 0.5$ for a very stable atmosphere. As low-altitude ($z<5$ m) wind speeds in a stable atmosphere are restricted to relatively low values (for higher wind speeds, stability is disrupted and the atmosphere becomes neutral), the spectral maximum may shift up to approximately $0.5V/z<3$ Hz. The inertial subrange thus expands or shrinks at its lower boundary, but its frequency dependency follows the "law of 5/3."

B. Effect on microphone in wind screen

The spectrum of longitudinal atmospheric turbulence in the inertial subrange was described in the previous section with the (squared) rms value of velocity variation per unit frequency $\mathbf{u}_f^2 = 0.3 \cdot u_*^2 (V/z)^{2/3} \cdot f^{-5/3}$. It is convenient to integrate this over a frequency range f_1–f_2 to obtain a 1/3-octave band level ($f_m = 2^{-1/6} \cdot f_2 = 2^{1/6} \cdot f_1$) with center frequency f_m: $\mathbf{u}_{1/3}^2 = 0.046 \cdot u_*^2 (f_m \cdot z/V)^{-2/3} = [0.215 \cdot u_* (f_m \cdot z/V)^{-1/3}]^2$. Substituting u_* from Eq. (6) and applying the result to Eq. (5) for 1/3 octave band levels $L_{\mathrm{at},1/3}(f_m) = 20 \log(\alpha \rho V \mathbf{u}_{1/3}/p_{\mathrm{ref}})$ yields

$$L_{\mathrm{at},1/3}(f) = 40 \log\left(\frac{V}{V_0}\right) - 6.67 \log\left(\frac{zf}{V}\right) - 20 \log\left[\ln\left(\frac{z}{z_0}\right) - \Psi\right] + C. \tag{8}$$

Here the frequency index m as well as the logarithm index 10 have been dropped, as will be done in the remainder of the text. In Eq. (8) $C = 20 \log(0.215 \kappa \alpha \rho V_0^2 / p_{\mathrm{ref}}) = 62.4$ dB for $\kappa = 0.4$, $\alpha = 0.25$, $\rho = 1.23$ kg/m^3 and pressure level is taken *re* $p_{\mathrm{ref}} = 20\ \mu$Pa. For octave band levels $L_{\mathrm{at},1/1}(f)$ the constant at the right-hand side of Eq. (8) is $C+4.8=67.2$ dB.

Equation (8) does not apply to frequencies where eddies are smaller than the wind screen. The contribution of small eddies to the pressure variation in the center of the screen will decrease proportionally to the ratio of eddy size ℓ^2 and wind screen surface πD^2 (where ℓ is the eddy length scale and $f \cdot \ell = V$). When this ratio decreases more eddies will simultaneously be present at the screen surface and resulting pressure fluctuations at the surface will more effectively cancel one another in the interior of the wind screen. The pressure variation in the wind screen center resulting from one eddy is proportional to the size of the eddy relative to the screen surface, i.e., ℓ^2/D^2, but also the screen center pressure resulting from the random contributions of all N eddies on the screen surface is proportional to $\sqrt{N}$, where $N \sim D^2/\ell^2$. The resulting screen center pressure is thus proportional to individual eddy pressure p_f and $(\ell^2/D^2)\sqrt{(D^2/\ell^2)} = \ell/D = V/fD$. Consequently a factor $-20 \log(fD/V)$ must be added to the resulting rms pressure level.

In wind noise reduction measured by Morgan[10] there is a change in frequency dependency at screen number $D/\ell \approx 1/3$ (also cited by Zheng and Tan).[3] So the frequency where, relative to Eq. (8), the decrease proportional to $20 \log(D/\ell)$ sets in, is $D/\ell = Df_c/V = 1/3$. As the change will be gradual, a smooth transition $-10 \log[1+(f/f_c)^2]$ can be added to Eq. (8). Writing $\log(zf/V)$ as $\log(fD/V) + \log(z/D)$ and $F(z) = -20 \log\{(z/D)^{1/3}[\ln(z/z_0) - \Psi]\}$, this yields

$$L_{\mathrm{at},1/3}(f) = 40 \log\left(\frac{V}{V_0}\right) - 6.67 \log\left(\frac{fD}{V}\right) - 10 \log\left[1 + \left(\frac{f}{f_c}\right)^2\right] + F(z) + C. \tag{9a}$$

The term before $F(z)$ is the equivalent of a first-order low pass filter with corner frequency $f_c = V/(3D)$. With usual screen diameters 5–25 cm and wind speeds 1–20 m/s, f_c is in the range of 1 to 100 Hz. With the common 10 cm diameter wind screen f_c will usually be in the infrasound region. $L_{\mathrm{at},1/3}$ according to Eq. (9a) has been plotted in Fig. 1 for different wind speeds and with $D=10$ cm, and $F(z) + C = 42$ dB. For $f < 0.5 f_c$ the term before C is less than 1 dB and Eq. (9a) reduces to Eq. (8).

Equation (9a) can be rewritten with Strouhal number $\mathrm{Sr} = fD/V$ as an independent variable of a "meteorologically reduced" 1/3 octave band level $L_{\mathrm{red},1/3} = L_{\mathrm{at},1/3} - 40 \log(V/V_0) - F(z)$

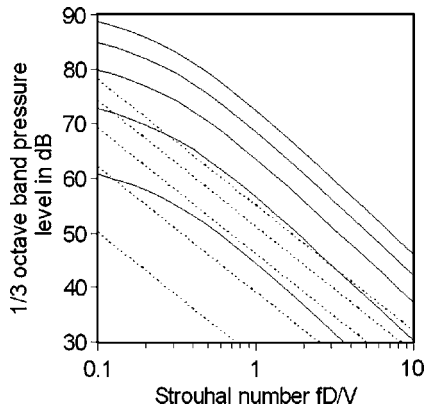


FIG. 1. Solid lines: 1/3 octave band pressure levels due to atmospheric turbulence on a screened microphone with $F(z)+C=42$ dB, $D=0.1$ m, at wind speeds of (bottom to top) 2, 4, 6, 8, and 10 m/s; dashed lines: pressure levels at same wind speeds according to Strasberg.

$$L_{\mathrm{red},1/3} = -6.67\log(\mathrm{Sr}) - 10\log[1+(3\mathrm{Sr})^2] + C. \quad (9b)$$

For frequencies $f \gg f_c$ the term before C in Eq. (9b) reduces to $-20\log(3\mathrm{Sr})$ and Eq. (9b) can be written as

$$L_{\mathrm{red},1/3} = -26.67\log(\mathrm{Sr}) + C - 9.5. \quad (10a)$$

This can be rewritten in aerodynamic terms, like Eq. (1), as

$$L_{p,1/3} = 20\log(\mathbf{p}_{1/3}/\rho V^2) = -26.67\log(\mathrm{Sr}) + F(z) + C_p, \quad (10b)$$

where $C_p = 20\log(0.215\kappa\alpha) - 9.5 = -43$ dB. For $F(z) = -20$ dB (e.g., a 10 cm diameter wind screen at a $z=2$ m, $z_0=5$ cm, and $\Psi=0$) the right-hand side of Eq. (10b) is $-26.67\log(\mathrm{Sr}) - 63$ dB. Comparison with Strasberg's result [Eq. (1) and gray lines in Fig. 1] shows that the frequency dependency is slightly different, and levels are 14–19 dB higher ($0.5 < \mathrm{Sr} < 10$), which is comparable to the difference in levels in the results of Boersma and Morgan *et al.* (see the Introduction). The change in slope, visible at Strouhal number $Df_c/V = 0.3$ in Fig. 1, is a feature not explained by the earlier authors.

C. Frequency regions

From the theory above it can now be concluded that the wind-induced pressure level on a (screened) microphone stretches over four successive frequency regions:

(i) At very low frequencies (less than a few hertz) the turbulence spectrum is in the energy-producing subrange; 1/3 octave band pressure level $L_{\mathrm{at},1/3}$ is independent of frequency (white noise), but increases with wind speed.

(ii) At frequencies up to $f_c = V/(3D)$, which is usually in the infrasound region, the turbulence spectrum is in the inertial subrange, $L_{\mathrm{at},1/3}$ is proportional to $-6.7\log f$ and $46.7\log V$.

(iii) At higher frequencies, but still in the inertial subrange, eddies average out over the wind screen more effectively at increasing frequency ($L_{\mathrm{at},1/3}$ proportional to $-26.7\log f$), but pressure level increases faster with wind speed ($L_{\mathrm{at},1/3}$ proportional to $66.7\log V$).

(iv) At frequencies beyond $0.1V/\eta_s$ (see Ref. 11, p. 586) atmospheric turbulence enters the dissipation range and turbulence vanishes. This is in the range $\mathrm{Sr} = fD/V > 0.1D/\eta_s \approx 100[D/\mathrm{m}] = D/\mathrm{cm}$.

The inertial subrange [(ii) and (iii)] is of most interest here, as it is within the commonly used range of acoustic frequencies and levels.

D. Wind-induced broadband A-weighted level

In Fig. 2 1/3-octave band levels according to Eq. (9) are plotted for different wind speeds for $F(z) = -20$ dB. Also, levels are plotted after A weighting to show the relevance to broadband acoustic measurements, where wind-induced noise may be a disturbance added to an A-weighted sound level. At the frequency where turbulent eddies enter the dissipation subrange ($f \approx 0.1V/\eta_s$), no further values are plotted as the turbulent velocity spectrum falls very steeply and induced pressure levels are considered negligible. A-weighted pressure levels $L_{\mathrm{at},A}$ can be calculated by summing over all 1/3-octave bands. The wind speed dependency can then be determined from the best fit of $L_{\mathrm{at},A}$ versus V

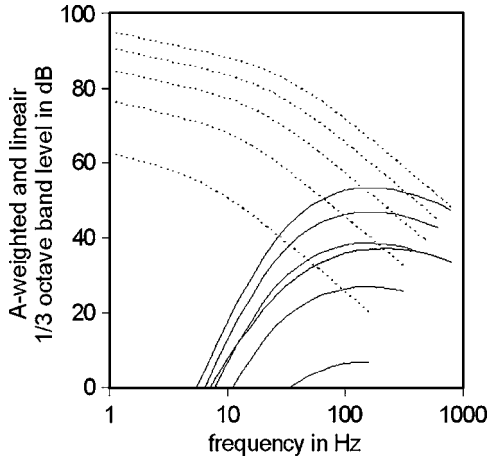


FIG. 2. Linear (dashed) and A-weighted (solid lines) 1/3-octave pressure levels due to atmospheric turbulence on a screened microphone with $F(z) + C = 42$ dB, $D = 0.1$ m, at wind speeds of (bottom to top) 2, 4, 6, 8, and 10 m/s; thick gray line: A-weighted pressure level according to Strasberg at 10 m/s.

$$L_{\mathrm{at},A} = 69.4\log(V/V_0) - 26.7\log(D/\ell_0) + F(z) + C - 74.8, \quad (11a)$$

where $\ell_0 = 1$ m is a reference length. Equation (11a) has the same structure as Eq. (10a), but a much smaller constant term as a result of A weighting. The V-related slope decreases somewhat with decreasing wind screen diameter and is 65.5 dB when $D = 1.25$ cm (unscreened $\frac{1}{2}$ in. microphone), but varies less than 1 dB for $5 < D/\mathrm{cm} < 50$. Equation (11a) is not very sensitive for the cutoff at $f = 0.1V/\eta_s$: if spectral levels according to Eq. (9) are integrated over all frequencies, total level does not increase significantly at high wind speeds, and with less than 3 dB at low wind speeds. It will be noted that the slope in Eq. (11a) with wind speed is slightly higher than for individual spectral levels for $f > f_c$ [66.7 dB, see Eq. (10a)], due to lower A weighting at the increasingly higher frequencies.

It is convenient to substitute $G(z) = F(z) - 6.7\log(D/\ell_0) + 14 = -20\log\{0.2(z/\ell_0)^{1/3}[\ln(z/z_0) - \Psi]\}$ as, for $z_0 = 2.5$–6 cm and $\Psi = 0$, $G(2\ \mathrm{m}) = 0 \pm 1$ dB, and the A-weighted, broadband pressure level caused by atmospheric turbulence from Eq. (11a) becomes

$$L_{\mathrm{at},A} = 69.4\log(V/V_0) - 20\log(10D/\ell_0) + G(z) + C - 68.8. \quad (11b)$$

The logarithm containing $10D$ is nil for the most widely used wind screen diameter. For a 10 cm diameter wind screen and measurement over a flat area with a low vegetation cover in neutral conditions $G(z) \approx 0$ and $L_{\mathrm{at},A} \approx 69.4\log(V/V_0) - 6.4$ dB(A).

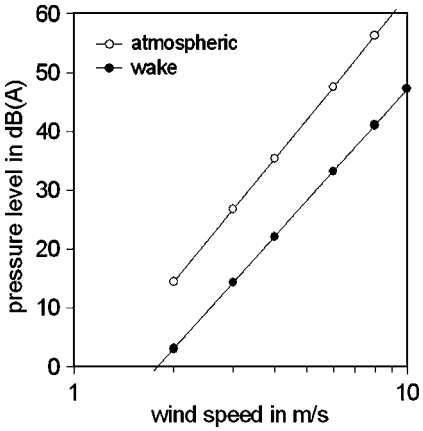


FIG. 3. A-weighted broadband pressure level caused by atmospheric or wake turbulence with $G(z)+C=62$ dB, $D=0.1$ m.

Figure 3 is a plot of Eq. (11) with $G(z)=0$, $C=62$ dB. Also plotted in Fig. 3 is the relation for wake turbulence according to Strasberg, obtained by A weighting and integrating Eq. (2) over frequency.

III. COMPARISON WITH EXPERIMENTAL RESULTS

A. Measured spectral pressure levels due to wind turbulence

Several authors have performed measurements to determine spectral levels due to wind, including wind-induced sound pressure fluctuations. Data are available from Larsson and Israelsson,[12] Jakobsen and Andersen,[13] and Boersma[4] and include screened as well as unscreened microphones. Table I gives an overview of measurement parameters. None of the authors give the degree of atmospheric stability, but for Jakobsen's data $\Psi \leqslant 0$ (night) and Boersma's $\Psi \geqslant 0$ (summer's day). Jakobsen mentions roughness height of the location (a golf course), Boersma grass height ($\approx$10 cm), Larsson only mentions measurement height over grass at either 1.25 or 4 m, without specifying which height applies to a measurement result. To prevent using spectra at large values of $|\Psi|$ no data at low wind speeds (<2 m/s at microphone) are used. This is also recommendable as at low wind speed sound not related to wind is more likely to dominate. Preferably L_{eq} values are used, but these are not available from Boersma who used 95th percentile levels (L_{95}). However, the L_{50} values are available from his original data. Though Boersma quotes $L_{eq} \approx L_{50}$, here $L_{eq} \approx L_{50}+3$ is preferred, as this is in agreement with long term data on wind noise,[14] and this is assumed to be valid for every frequency band.

Also, L_{eq} values are presented from measurements made by the author at several locations; at one location (Zernike) for the purpose of wind noise measurements, and otherwise (Horsterwold, Kwelder) selected for having little other noise. Here also the degree of atmospheric stability is unknown, as at the time of measurement it was not known to be a relevant factor. The "Zernike" measurements were done at the university grounds (latitude 53°14′43″, longitude 6°31′48″) with both the microphone (in a spherical foam screen of 2.5, 3.8, or 9.5 cm diameter) and the wind meter at 1.2 or 2.5 m over grass at least several hundred meters from trees, and an estimated roughness height of 5 cm. They were performed in daytime in December 2003 and August 2004 with a fair wind under heavy clouding. The "Kwelder" measurements were made in daytime or evening in July and August of 1996 at an open area at the Dutch coast (latitude 53°25′46″, longitude 6°32′40″), consisting of level land overgrown with grass and low weeds and close to tidal water. Sound measurements were taken at a height of 1.5 m at times when no sound could be heard but wind-related sound and distant birds. The microphone was fitted with a spherical 9.5 cm diameter foam

TABLE I. Measurement characteristics.

Author	Period	Location	z_0 (cm)	H_{wind} (m)	H_{mic} (m)	V_{mic} (m/s)	D (cm)	T (min.)	N[a]	F (Hz)	Band width[f]
Larsson *et al.*	Late summer to early autumn	Grass lawn	5[b]	mic	1.25 or 4	2–7	no[d] 9.5	6 obs.[e]	9 9	63–8000	1/1
Jakobsen *et al.*	Summer to Dec, night	Golf course	2	10	1.5	3–7	9.5/25	?[e]	10	63–8000	1/1
Boersma	Summer, day	Grassland	3[b]	2	1.5	3–7 2–9	no[d] 9	160 430	9 7	6–16000 6–16000	1/3
This study:											
Horsterwold	Clouded winter night	Grass, reeds, trees	60[c]	10	2	4–6	9.5	230	4	31–8000	1/1
Kwelder	Clear summer day,	Grass, herbs	2[b]	5	1.5	3–5	9.5	40	6	6–16000	1/3
Zernike	Clouded summer day	Grassland	5[b]	1.5	2.5	5	2.4/3.8/9.5	30	3	6–1000	1/3
	Clouded winter day				1.2	4	3.8/9.5	20	2	1–1000	

[a]Number of measurements.
[b]Estimated.
[c]Fitted.
[d]No=unscreened.
[e]Observations of unknown length.
[f]1/1 or 1/3 octave band.

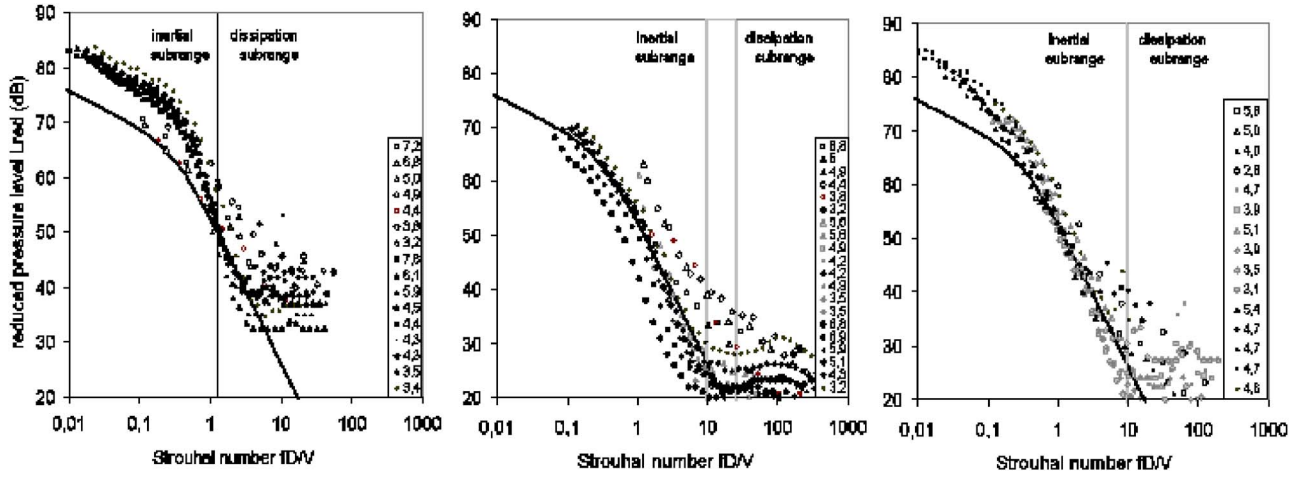


FIG. 4. Reduced 1/3 octave band pressure levels at different wind speeds (in legend: m/s), bold line is predicted spectrum. Left: unscreened microphone, from Larsson *et al.* (open symbols) and Boersma (black); middle: screened microphone, from Larsson *et al.* (open symbols), Jakobsen *et al.* (gray), and Boersma (black); right: screened microphone, measurements in Horsterwold (open symbols), Kwelder (gray) and Zernike (black).

wind screen. Wind speed at microphone height at 1.5 m was estimated from measured wind speed at 5 m height with Eq. (6), z_0 estimated as 2 cm. Finally the "Horsterwold" measurements were made in December 2001 in an open space with grass and reeds (latitude 52°18′3″, longitude 5°29′38″) between 5 and 10 m high trees at a distance of approximately 30 m but further in the windward direction, in a mostly clouded night. Wind speed and sound were measured at 2 m height, the wind screen was a 9 cm diameter foam cylinder. Due to the differences in vegetation, roughness length here was difficult to estimate, and was determined by fitting measurement results to the expected level (resulting in 60 cm and a more limited range of values of Ψ to fit).

At very low frequencies in our Zernike measurements the 1/3-octave band levels were corrected for reduced response. The frequency response of the B&K $\frac{1}{2}$ in. microphone type 4189 is specified by Brüel and Kjaer[15] and is effectively a high pass filter with a corner frequency of 2.6 Hz. The response of the Larson Davis type 2800 frequency analyzer is flat (±1 dB) for all frequencies.

To plot spectra the reduced pressure level $L_{\mathrm{red},1/3}$ is calculated, leaving only the screen diameter based Strouhal number $\mathrm{Sr}=fD/V$ as the independent variable. Octave band pressure levels $L_{\mathrm{red},1/1}$ are substituted by $L_{\mathrm{red},1/3}+4.8$. As atmospheric stability is as yet unknown, the stability function is set to zero. If wind velocity was not measured at microphone height, the logarithmic wind profile [Eq. (6) with $\Psi=0$] is used to determine V_{mic} from the wind speed at height z.

Unweighted spectra of 1/3-octave levels are plotted in the left part of Fig. 4 for the unscreened microphones from data from Larsson and Boersma. Also plotted is the spectrum according to Eq. (9b), valid for the inertial subrange. Due to the small size of the unscreened microphone (1.25 cm) part of the spectrum lies in the dissipation range at frequencies $f>0.1V/\eta_s\approx 100$ V/m, corresponding to $\mathrm{Sr}>100$ D/m $=1.25$.

In Fig. 4(b) spectra are plotted from screened microphones, from data from Larsson, Jakobsen, and Boersma. As these spectra were determined with a range of screen diameters, the change from the inertial to the dissipation subrange extends over a range of non-dimensional frequencies (Strouhal numbers). Finally Fig. 4(c) shows spectra from the Horsterwold, Zernike, and Kwelder measurements. In all fig-

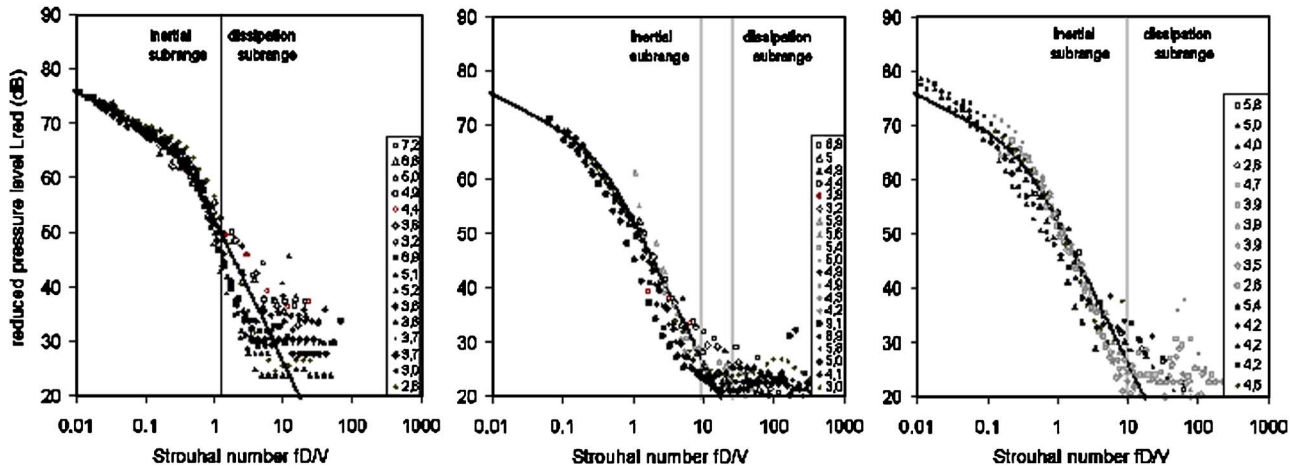


FIG. 5. Same as Fig. 4, but after fitting with stability function Ψ.

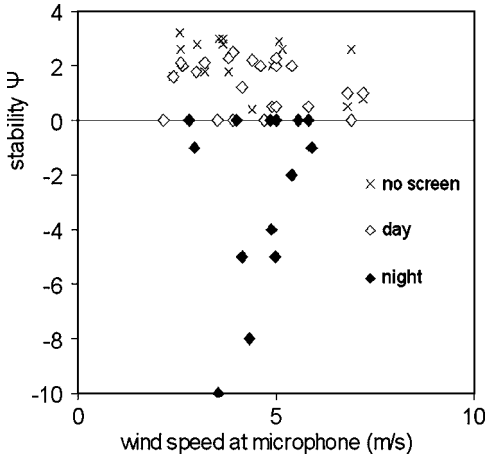


FIG. 6. Values of stability function Ψ after fitting reduced spectral levels with theoretical spectrum, for measurements in day or night time, and for unscreened microphones in daytime.

ures spectra deviate from the predicted spectrum at high Strouhal numbers because either the lower measurement range of the sound level meter is reached or ambient sound dominates the wind-induced pressure level. Also, at these high Strouhal numbers most values are in the dissipation range where the present model is not valid.

In Fig. 4 atmospheric stability has not been taken into account yet (in fact $\Psi=0$ was used), due to lack of data to determine Ψ. In stable conditions ($\Psi<0$) L_{red} will be higher, in unstable conditions ($\Psi>0$) lower, causing the plotted spectra to shift vertically if the proper value $\Psi\neq 0$ is applied.

If wind speed at microphone height is deduced from wind speed at another height, the shift is more complex, as stability then also affects the term $40\log(V/V_0)$ as well as the ordinate value $\mathrm{Sr}=fD/V$. The approach taken here is to vary Ψ to obtain a best fit to the theoretical value of the L_{red} levels at nondimensional frequencies in the inertial subrange. The fitted spectra are plotted in Fig. 5. The values of Ψ that gave the best fits are plotted in Fig. 6, categorized in daytime and nighttime measurements (where one would expect $\Psi\geqslant 0$ and $\Psi\leqslant 0$, respectively). Measurements with unscreened microphones are indicated separately, and are in daytime for Boersma's measurements and probably also for Larsson's, so one would expect $\Psi\geqslant 0$.

B. Measured broadband pressure levels due to wind turbulence

Several authors give a relation between broadband A-weighted sound pressure level L_A and wind speed V.[4,10,11] According to Boersma $L_A\sim 22.6\log(V)$ (with V measured at 2 m height, L_A at 1.5 m), to Larsson $L_A=4.4\cdot V+27.5$ (V and L_A measured at the same height of either 1.5 or 4 m), to Jakobsen $L_A=6.8\cdot V-2.6$ (V measured at 10 m, L_A at 1.5 m). However, as Boersma clearly shows, most of the A-weighted sound is due to ambient wind-induced sound, especially at low wind speeds. So these relations cannot be used to describe sound levels induced by wind on the microphone.

A practical situation where the influence of wind on the microphone+wind screen could be investigated directly offered itself when on May 28, 2000 a storm occurred during our "Wieringerwaard" measurements. The microphone, in a 9 cm foam cylinder, and a wind meter were both placed at a height of 4.6 m, 2 m apart, in front of a big farmer's shed 5 m to the west of the microphone (latitude 52°48′41″, longitude 4°52′23″). A second, "free wind" wind meter at 10 m height was placed further away to measure undisturbed wind. Around the measurement location were fields with potato plants of 20—30 cm height. As it was May, an unstable atmosphere is expected in daytime, leaning to neutral when the wind speed increases.

Some measurement results are given in Fig. 7 (plotted values are 10 min averages of samples measured at a rate of 1 s^{-1}). In the left part of the figure the "free" wind speed v_{10} is seen to increase to 20 m/s in the course of the day after a relatively quiet night. The wind speed V_{mic} near the microphone increased at practically the same rate between 6 and 12 o'clock, but then abruptly fell from 13 to 2 m/s and thereafter remained at a low value even while the "free" wind speed was still increasing. Up to 12 o'clock the sound level (equivalent A-weighted level per 10 mins) increased in proportion to the wind speed reaching a maximum of 84 dB(A), but then fell abruptly to 50 dB(A) at the same time the wind speed at the microphone collapsed. In this morning the unobstructed wind began in the east and gradually turned south. When at 12 o'clock the wind passed behind the shed, the microphone was suddenly taken out of the wind. There is no reason that the *sound* reaching the micro-

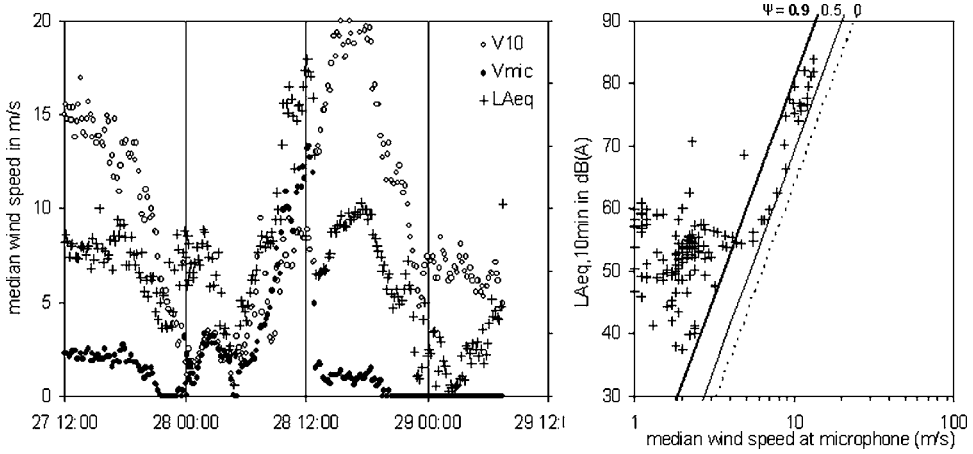


FIG. 7. Measurements during a storm on May 28, 2000 in front of a big shed; left: 10 min averages of wind speeds at microphone height and 10 m height and equivalent sound pressure level L_{Aeq}; right: L_{Aeq} as a function of microphone wind speed and predicted sound pressure level for three values of Ψ.

phone changed significantly, but due to the sudden wind speed reduction the measured sound pressure level dropped by 34 dB(A). After that the sound pressure level increased again up to 60 dB(A) as the storm was still gaining strength, producing more ambient sound. The measured pressure level above 60 dB(A) is wind-induced "pseudo"- sound, that is, sound resulting from moving air, not from airborne sound.

In the right part of Fig. 7 the A-weighted equivalent (pseudo-) sound pressure level per 10 mins over the same period as in the left part of Fig. 7, is plotted as a function of wind speed at the microphone. There is an obvious direct correlation between pressure level and wind speed at higher wind speeds ($V \geqslant 6$ m/s) in contrast to the levels at lower wind speeds. Again, the stability factor Ψ is not known, but in daytime and in strong winds it must be small and positive. In Fig. 7 results are shown for two values of Ψ (with $z_0 = 20$ cm) encompassing the measured values.

C. Screen reduction

For two of our Zernike summer measurements (see Table I) with place and atmospheric conditions (near neutral, $V_{\mathrm{mic}}=5.0$ m/s) unchanged within the measurement period, the difference between 1/3-octave band pressure levels measured with an approximately spherical 2.4 cm wind screen and a spherical 9.5 cm wind screen are plotted in Fig. 8. Also plotted is the calculated difference in screening effect based on Eq. (9a), with only both terms before C differing between both measurements. It appears that the measured difference in screening effect is on average approximately 1 dB higher than the calculated level. It is not clear why the difference in screening is negative at frequencies below 2 Hz. For a somewhat smaller ($1.8\ \mathrm{cm} < D < 2.4\ \mathrm{cm}$) wind screen the difference would agree better with the calculated effect.

IV. DISCUSSION

The model developed in this paper starts with the assumption that wind-induced "sound" pressure levels on a microphone are caused by atmospheric turbulence. Then, at low non-dimensional frequencies ($\mathrm{Sr} \ll 0.3$) spectral levels are determined entirely by atmospheric turbulence. In this frequency range a wind screen has no effect. At higher frequencies, where pressure fluctuations tend to cancel one another more effectively as their scale decreases relative to the wind screen diameter, a wind screen acts as a first-order low pass filter for turbulent fluctuations. In this frequency range ($\mathrm{Sr} > 0.3$) a wind screen diminishes the effect of turbulence, and better so if it is bigger.

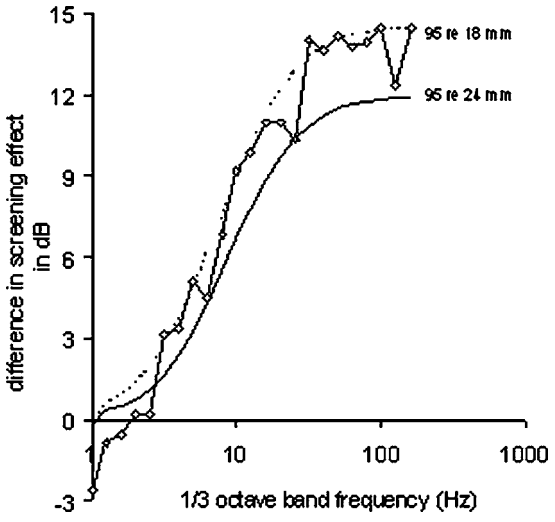


FIG. 8. Measured (line with markers) and calculated screening effect of a 9.5 cm relative to a 2.4 or 1.8 cm diameter wind screen.

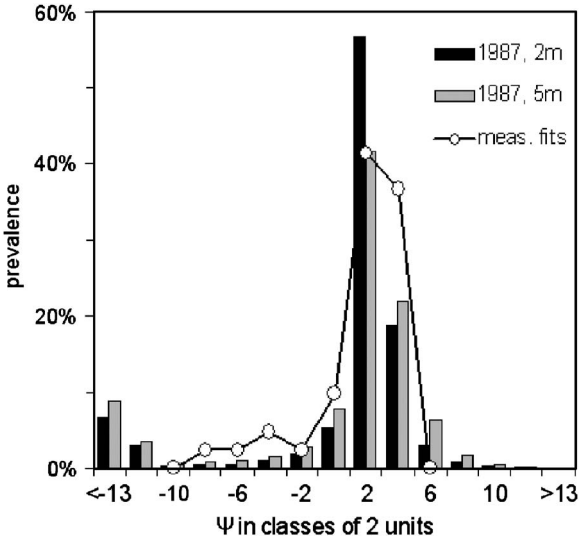


FIG. 9. Frequency distributions of stability factor Ψ at 2 and 5 m height, based on $\frac{1}{2}$ h observations over 1987 (bars) and resulting from fitted spectra (open circles); class x: $x-1 < \Psi < x+1$.

Wind-induced pressure levels are determined not just by wind speed and screen diameter, but also by two factors that are relevant for the production of turbulence: atmospheric instability, and surface roughness. The stability factor Ψ and roughness height z_0 are determinants for thermal and frictional turbulence, respectively. These determinants are usually not taken into account with respect to wind-induced noise and are consequently not reported. Atmospheric stability therefore had to be estimated by varying the value of Ψ until a best fit was obtained of measured spectra to the calculated spectrum. Roughness length, when unknown, was assumed to be comparable to vegetation height.

The values of Ψ that resulted in the best fits are shown in Fig. 6. They can also be compared to values obtained from long term measurements at the Cabauw measurement site of the Royal Netherlands Meteorological Institute (KNMI). The Cabauw site is in open, flat land west of the central part of the Netherlands and may be considered representative for locations in comparable terrain in the north and central parts of the Netherlands (Boersma's and our measurements), Denmark (Jakobsen *et al.*) and the Swedish Uppsala plain (Larsson *et al.*). KNMI provided us with a data file containing 30 min averages of the Monin-Obukhov length L over one year (1987). From this the dimensionless height $\zeta = z/L$ can be calculated and then the stability factor Ψ [see text below Eq. (6)]. In Fig. 9 the frequency distribution is shown of all 17 520 ($=2 \cdot 24 \cdot 365$) values of Ψ, for two altitudes: 2 and 5 m. Also the frequency distribution is shown of the 42 values of Ψ resulting from our fitting procedure. The distribu-

tion of our fitted values resemble the distribution of actually occurring values (in 1987) and thus seems plausible.

Two constants are not known accurately: the constant α, assumed to have a value 0.25, and the ratio of screen diameter and eddy size at the corner frequency, where 3 was used. If the slopes are as in Eq. (9b), the best fit of all data points in Fig. 5 at $\mathrm{Sr}<2.5$ is a line $L_{\mathrm{red},1/3}=-6.7\log(\mathrm{Sr})-10\log[1+(3.8\cdot\mathrm{Sr})^2]+62.0$. So the ratio ℓ/D (3.8) where screen averaging over eddies sets in may be greater than assumed (3); the constant term may be somewhat smaller, which could be a result of a lower value of α than assumed (0.24 instead of 0.25). The fit is within 2.2 dB of the calculated value [Eq. (9b)]; for $2.5<\mathrm{Sr}<16$ this fit is on average 2.1 dB above the calculated value. The standard deviation of the measured 1/3 Strouhal octave band levels is less than 3.5 dB at $\mathrm{Sr}<2.5$ and up to 7 dB at $2.5<\mathrm{Sr}<16$.

V. APPLICATIONS

As microphone wind noise appears to be closely correlated to atmospheric turbulence, acoustic measurements can alternatively be used to measure turbulence spectra or turbulence strength, especially in the inertial subrange. This provides a new way to determine (e.g.) friction velocity or atmospheric stability. As the measured signal decreases above the corner frequency $f_c=V/(3D)$ this frequency is best chosen high, which can be achieved with a small, bare microphone.

The present model can be used to distinguish wind-induced noise from other wind related sound. An application is the measurement of wind turbine sound or (without an operating wind turbine) ambient background sound in relatively strong winds. If the measurement is on a wind exposed site it is probable that at high wind speeds wind induced noise influences or even dominates either wind turbine sound or proper ambient sound. A measured level can now be corrected for wind-induced sound with a calculated wind noise level. In less exposed sites it is usually not clear in what degree the measured levels are influenced by wind-induced noise. To calculate wind-induced noise levels additional measurements are necessary to determine roughness height and atmospheric stability. Stability can be estimated from wind speed measurements on two heights, using Eq. (6). Roughness height can be estimated from tabulated values or from wind speed measurement at two heights in a neutral atmosphere, at times when the logarithmic wind profile is valid [Eq. (6) with $\Psi=0$]. In neutral and stable conditions wind-induced noise levels are not very sensitive to errors in roughness height: With an error of a factor of 2 in $z_0=10$ cm, the level changes less than 2 dB if microphone height is 3 m or more.

VI. CONCLUSION

Measured spectra, reduced with a term for wind speed and turbulence strength, coincide well with calculated values for unscreened as well as screened microphones in the range where the theoretical model [Eq. (9)] is valid. To test the model more thoroughly, measurements should include a determination of roughness length and atmospheric stability, in addition to the usual measurement of wind speed and measurement height.

The model shows that to avoid high wind-induced pressure levels, it is best to measure at low wind speed and with a large diameter wind screen, which is common knowledge in acoustics. The overall reduction ΔL_A from a bigger wind screen relative to a smaller one is determined by the ratio of the screen diameters D_1 and D_2: $\Delta L_A=20\log(D_2/D_1)$ [from Eq. (11b), $D>5$ cm]. A wind screen does not reduce noise from atmospheric turbulence at frequencies $f<V/(3D)$.

The model also shows that to reduce wind induced sound, it helps to measure over a low roughness surface and at night (stable atmosphere), as both factors help to reduce turbulence, even if the (average) wind speed on the microphone does not change. With reduced turbulence, wind-induced pressure levels will finally reach the level given by Strasberg [Eqs. (1) or (2)], where turbulence is the result of the wake caused by the wind screen.

One might be tempted to think that a higher measurement altitude would also help to reduce wind noise, as this would make $G(z)$ in Eq. (11b) more negative, thus reducing $L_{\mathrm{at},A}$. However, increasing altitude will in practice lead to higher wind speeds, especially so in a stable atmosphere, and the first term in Eq. (11b) would more than compensate the decrease in $G(z)$. It is therefore preferable to measure at low altitude if less wind noise is desired.

List of Symbols

α = constant [Eq. (3)]
ΔL_A = reduction in broadband A-weighted wind noise level due to a wind screen
κ = von Karman's constant=0.4
η_s = Kolmogorov size$\approx$1 mm
λ = wavelength
ν = kinematic viscosity of air
ρ = air density
Ψ = stability function=$\Psi(\zeta)$
ζ = dimensionless height=z/L
c = velocity of sound in air
C = air density dependent constant $=20\log(0.215\kappa\alpha\rho V_0^2/p_{\mathrm{ref}})=62.4$ dB [Eq. (8)]
C_p = constant$=20\log(0.215\kappa\alpha)-9.5=-43$ dB [Eq. (10b)]
D = wind screen diameter
f = frequency
f_m = middle frequency of 1/3 octave band
f_c = corner frequency for wind screen attenuation $=V/(3D)$
$F(z)$ = turbulence function$=-20\log\{(z/D)^{1/3}[\ln(z/z_0)-\Psi]\}$ [Eq. (9a)]
$G(z)$ = adapted turbulence function$=-20\log\{0.2(z/\ell_0)^{1/3}[\ln(z/z_0)-\Psi]\}$ [Eq. (11b)]
k = exponent in relation pressure-wind speed [Eq. (3)]
ℓ = turbulent eddy size
ℓ_0 = reference length=1 m

L = Monin Obukhov length [below Eq. (6)]
L_{eq} = equivalent sound pressure level re p_{ref}
L_{50} = 50th percentile of or median sound pressure level re p_{ref}
L_{95} = 95th percentile of sound pressure level re p_{ref}
$L_{1/3}$ = 1/3 octave band pressure level due to wake turbulence re p_{ref} [Eq. (2)]
L_{at} = pressure level due to atmospheric turbulence re p_{ref} [Eq. (5)]
$L_{at,A}$ = broadband A-weighted pressure level due to atmospheric turbulence re p_{ref}
L_A = broadband A-weighted sound pressure level re p_{ref}
$L_{p,1/3}$ = 1/3 octave band pressure level re ρV^2 [Eq. (1)]
$L_{red,1/3}$ = "meteorologically reduced" 1/3 octave band level pressure level due to atmospheric turbulence re p_{ref} [Eq. (9b)]
M = Mach number$=V/c$
n = nondimensional requency$=fz/V$
p = turbulence associated time varying part of pressure$=P-P_{average}$
P = air pressure
$P_{average}$ = time-averaged air pressure
p_f = pressure in relatively narrow frequency band centerd at frequency f
p_{ref} = reference sound pressure$=20$ μPa
Re = Reynolds number$=DV/\nu$
Sr = Strouhal number$=fD/V$
u = horizontal in-line component of variable part of wind speed$=U-V$
U = wind speed
u_* = friction velocity [Eq. (6)]
$\mathbf{u}_f$ = horizontal rms wind velocity variation per unit frequency
v = horizontal cross-wind component of variable part of wind speed
V = time averaged wind speed or speed in air
V_0 = reference velocity$=1$ m/s
V_{mic} = average wind speed at microphone height
w = vertical component of variable part of wind speed
z = height
z_0 = roughness height

Subscripts

1/1 = per octave band
1/3 = per 1/3 octave band
at = atmospheric turbulence

[1]M. Strasberg, "Dimensional analysis and windscreen noise," J. Acoust. Soc. Am. **83**, 544–548 (1988).
[2]S. Morgan and R. Raspet, "Investigation of the mechanisms of low-frequency wind noise generation outdoors," J. Acoust. Soc. Am. **92**, 1180–1183 (1992).
[3]Z. C. Zheng and B. K. Tan, "Reynolds number effects on flow/acoustic mechanisms in spherical windscreens," J. Acoust. Soc. Am. **113**, 161–166 (2003).
[4]H. F. Boersma, "Characterization of the natural ambient sound environment: measurements in open agricultural grassland," J. Acoust. Soc. Am. **101**, 2104–2110 (1997).
[5]N. O. Jensen and N. E. Busch, "Atmospheric turbulence," in *Engineering Meteorology*, edited by E. J. Plate (Elsevier Scientific Publishing Company, Amsterdam, 1982), pp. 179–229.
[6]H. Zhang, J. Chen, and S. Park, "Turbulence structure in unstable conditions over various surfaces," Boundary-Layer Meteorol. **100**, 243–261 (2001).
[7]D. K. Wilson and D. W. Thomson, "Acoustic propagation through anisotropic, surface-layer turbulence," J. Acoust. Soc. Am. **96**, 1080–1095 (1994).
[8]J. R. Garrat, *The Atmospheric Boundary Layer* (Cambridge University Press, Cambridge, 1992).
[9]J. C. Kaimal, J. C. Wyngaard, Y. Izumi, and O. R. Coté, "Spectral characteristics of surface-layer turbulence," Q. J. R. Meteorol. Soc. **98**, 563–589 (1972).
[10]M. S. Morgan, "An investigation of the sources and attenuation of wind noise in measurement microphones," Dissertation, University of Mississippi, 1993.
[11]E. J. Plate, "Wind tunnel modelling of wind effects in engineering," in *Engineering Meteorology*, edited by E. J. Plate (Elsevier Scientific Publishing Company, Amsterdam, 1982), pp. 573–639.
[12]C. Larsson and S. Israelsson, "The effects of meteorological parameters on sound propagation from a point source," Meteorologiska Institutionen Report No. 67, Kungliga Universitet Uppsala, Uppsala (1982).
[13]J. Jakobsen and B. Andersen, "Wind noise. Measurements of wind-generated noise from vegetation and microphone system," Danish Acoustical Institute Report No. 108, Lyngby (1983).
[14]G. P. van den Berg, *Statistics of Wind-Related Sound in Outdoor Monitoring* (Internoise, Prague, 2004).
[15]B&K, *Microphone Handbook* (Brüel & Kjaer, Denmark, 1995).
[16]M. Strasberg, "Nonacoustic noise interference in measurements of infrasonic ambient noise," J. Acoust. Soc. Am. **66**, 1487–1493 (1979).

Framework for wind noise studies

Richard Raspet, Jeremy Webster,[a] and Kevin Dillion
Department of Physics and Astronomy and the National Center for Physical Acoustics, University of Mississippi, Box 1848, University, Mississippi 38677

(Received 24 February 2005; revised 3 November 2005; accepted 9 November 2005)

Morgan and Raspet [J. Acoust. Soc. Am. **92**, 1180–1183 (1992)] performed simultaneous wind velocity and wind noise measurements and determined that the wind noise spectrum is highly correlated with the wind velocity spectrum. In this paper, two methods are developed for predicting the upper limits of wind noise pressure spectra from fluctuating velocity spectra in the inertial range. Lower limits on wind noise are estimated from two theories of the pressure fluctuations that occur in turbulence when no wind screen or microphone is present. Empirical results for the self-noise of spherical and cylindrical windscreens in substantially nonturbulent flows are also presented. Measurements of the wind velocity spectra and wind noise spectra from a variety of windscreens are described and compared to the theoretical predictions. The wind noise data taken at the height of the anemometer lies between the upper and lower limits and the predicted self-noise is negligible. The theoretical framework allows windscreen reduction to be evaluated in terms of the turbulent inflow properties and establishes practical upper limits on wind noise reduction for varying wind conditions. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146113]

PACS number(s): 43.28.Ra, 43.28.We, 43.28.Gq [DKW] Pages: 834–843

I. INTRODUCTION

Research on wind noise and wind screens for outdoor measurements has relied largely on comparison experiments. Measurements of wind noise from windscreened and bare microphones are compared to determine the wind noise reduction. Morgan and Raspet[1] used simultaneous wind noise and wind velocity measurements to show that the principle source of wind noise in a bare microphone outdoors is the interaction of the incoming velocity fluctuations with the microphone.

Subsequent research in the use of active cancellation between hot wire velocity sensors and microphones to reduce wind noise has verified the high correlation between wind velocity fluctuations and wind noise in a bare microphone.[2,3] In addition, velocity measurement devices have been constructed that measure the pressure distribution around a sphere and infer the fluctuating velocity from that pressure distribution.[4,5]

The wind velocity turbulence spectrum is heavily weighted to low frequencies; therefore, the wind noise levels are higher at low frequencies. This dependency conflicts with the desire to exploit the low frequency portion of the source spectrum in detection/identification systems since low frequency sound is less affected by atmospheric attenuation and diffraction than higher frequencies.

In this paper we primarily study wind noise prediction and generation for compact transducer systems mounted above the ground. A microphone system is considered compact for a given turbulence frequency or wave number if the turbulence scale is larger than the system. At higher frequencies, spatial averaging occurs around the windscreen and the wind noise is no longer correlated with the incoming turbulence.

Turbulence spectra are commonly divided into three regions: the source region, the inertial range and the dissipation region.[6] In the inertial range the power spectral density has a frequency or wave number dependence of $f^{-5/3}$ or $k^{-5/3}$, where f is the frequency and k the wave number. For measurements in windy conditions with microphones mounted at heights on the order of a meter, the lower limit of the inertial range will correspond to a few Hertz, and the upper limit to around 10^4 Hz. In this paper we will fit velocity spectral data to the five-thirds power law in the wave number range from 1.0 m^{-1} to 30.0 m^{-1} and use the fits to predict pressure contributions. The theoretical analysis is significantly simplified for power law spectra. In principle, this analysis can be extended into the source region for studies of infrasound.

The first set of calculations investigates the maximum pressure fluctuations measured on a bluff body in turbulent flow outdoors. In aerodynamic studies, it is common practice to normalize drag, lift and pressure data by the stagnation pressure,

$$p_s = \tfrac{1}{2}\rho V^2, \tag{1}$$

where V is the flow speed. The stagnation pressure is measured at the front of the body and is the largest positive pressure on its surface. We suggest using the fluctuating stagnation pressure calculated from the measured fluctuating wind velocity as the reference level for wind noise reduction. We will show that wind noise from a bare microphone is well approximated by such a calculation. The calculated stagnation pressure should represent an upper bound for the wind noise measured by microphones within aerodynamically bluff bodies in turbulent flows. Two methods for cal-

[a]Electronic mail: jwebster@olemiss.edu

culating the fluctuating stagnation pressure will be developed in Sec. III.

Windscreens are used to reduce wind noise in microphones. Spherically shaped ones are particularly effective. The basic reduction mechanism at low frequencies is thought to be that a microphone at the center of the sphere measures the area-averaged fluctuating pressures around the sphere. At low frequencies, the pressure fluctuations at the front of the sphere are opposite in sign from the pressure fluctuations at the rear for a given velocity fluctuation, and cancellation occurs between these contributions.[1,7–9] At high frequencies, when the diameter of the windscreen is larger than the turbulence scale, the velocity fluctuations are uncorrelated around the sphere and pressure averaging leads to much higher reductions (up to 35 dB).

The wind noise reduction below the fluctuating stagnation pressure level may be limited by other contributions to the pressure. For example, Schomer *et al.*[10] performed measurements that combined a two microphone processing system and a large windscreen system. The two microphone system, which consisted of two bare microphones separated vertically by 1.1 m, produced a C-weighted wind noise reduction of 33 dB while the passive windscreen system alone produced a reduction of 30 dB. The combination of the two systems, however, resulted in only a 33 dB reduction. This limit may be determined by the turbulent pressure fluctuations that exist regardless of the presence of a microphone in the flow. Morgan[1,9] refers to these fluctuations as "intrinsic pressure fluctuations." The prediction and measurement of the turbulent pressure fluctuations have been the focus of many fluid dynamic and meteorological investigations.[11–15] Predictions for turbulent pressure fluctuations due to turbulence-turbulence interaction, turbulence-shear interaction, and buoyancy have been developed.[12,13] In a large-eddy simulation (LES) study[13] and in turbulent free shear flow measurements,[11] the largest contribution to the total fluctuating pressure in the inertial range was the turbulence-turbulence interaction. In this paper, it is assumed that the turbulence-turbulence interaction is the dominant intrinsic fluctuation mechanism. In Sec. II, two theories for predicting the turbulence-turbulence interaction pressure from measured velocity fluctuation spectra will be described. This analysis will also be restricted to the inertial range. The turbulent pressure fluctuation defines an upper limit for wind noise reduction in the turbulent atmosphere and can be used to predict the maximum wind noise reduction possible with a compact sensor (without multisensor averaging).

Another limit on wind noise measurements is the self noise produced by the windscreen in the flow. Strasberg[16,17] has collected data for wind noise measured inside spherical and cylindrical porous windscreens in substantially nonturbulent flow, and demonstrated that the nondimensional data can be fit with a single curve. The results of Strasberg's research are described in Sec. IV.

The description of an experiment that simultaneously measured the three-dimensional wind velocity outdoors and the pressure fluctuations using a variety of pressure measuring sensors, microphones and windscreens is presented in Sec. V. F. Douglas Shields made these measurements as part of an investigation of wind noise relative to a unimorph pressure sensing system. The experimental data is provided to illustrate the use of the theoretical predictions in interpretation and analysis of experimental investigations of wind noise. The velocity and pressure spectra are also presented in that section.

Section VI compares predictions of stagnation pressure spectra and the turbulent pressure fluctuation spectra based on measured turbulent velocity spectra and compares the prediction to wind noise pressure spectra measured using a variety of microphones, sensors and windscreens and presents the conclusions of the study.

II. PRESSURE FLUCTUATIONS IN TURBULENT FLOWS

In this section, two related theories are developed to estimate the fluctuating pressure spectrum from the measured turbulence velocity spectrum in the inertial range. George, Beuther, and Arendt[11] provide a good review of the research and theoretical developments up to 1984. The emphasis of George *et al.* is on the measurement and identification of the different pressure fluctuation sources from a turbulent jet without interference from wind noise on their streamlined pressure probe. They identified two potential sources of intrinsic pressure fluctuation in the flow: turbulence-turbulence interaction and turbulence-mean shear interaction. For atmospheric flows away from the ground surface, we expect the mean shear to be small compared to the free jet studied in Ref. 11.

For the comparison of ideal predicted pressure fluctuations with measurement, we follow Batchelor's calculation[14] to relate the measured velocity fluctuation spectrum in the inertial region to the predicted pressure fluctuation spectrum. The second method we use to predict the pressure spectrum in the inertial range is due to Miles, Wyngaard, and Otte.[13] They calculate velocity and pressure fluctuation spectra from an LES and fit their calculated data to scaling laws based on Obukhov's work.

A. Batchelor's theory for fluctuations in isotropic homogeneous flows

Batchelor[14] begins with the pressure equation for incompressible flows,

$$\frac{1}{\rho}\nabla^2 p = -\frac{\partial^2 u_i u_j}{\partial x_i \partial x_j}, \tag{2}$$

where p is the pressure fluctuation, ρ is the ambient density, and u_i and u_j are velocity components. Summation notation is used. He assumes an isotropic, homogeneous turbulence field and examines the spatial correlations of Eq. (2) at two different positions to develop relations between the structure functions for the velocity and pressure. In order to solve the resulting equation, a further assumption is imposed that the fourth moments of velocity are related to the second moments as if u_i and u_j are normally distributed. This also results in third moments being zero. The second assumption is partially based on measurements and partially based on ne-

cessity in the absence of data or theory applicable to the third and fourth moments.

The result relevant to this study relates the longitudinal correlation function of the fluctuating velocity component in the average flow direction to the pressure correlation in the direction of the mean velocity. In an isotropic homogeneous flow, all terms in the velocity correlation tensor, $R_{ij}(\mathbf{r})$, can be expressed in terms of a single scalar $f(r)$ such that

$$R_{11}(r,0,0) = \overline{u^2} f(r). \tag{3}$$

In the inertial range, the velocity structure function is given by

$$\overline{(u-u')^2} = 2\overline{u^2}(1-f) = C(\nu\epsilon)^{1/2}\left(\frac{r}{\eta}\right)^{2/3}, \tag{4}$$

where u and u' are the velocity components measured longitudinal to the displacement r at x and $x+r$, respectively, $\eta = \nu^{3/4}\epsilon^{-(1/4)}$, ν is the kinematic viscosity, ϵ is the mean energy dissipation per unit mass of fluid, and C is an arbitrary constant. The equivalent result developed by Batchelor[14] for pressure is

$$\frac{1}{\rho^2}\overline{(p-p')^2} = C^2\epsilon\nu\left(\frac{r}{\eta}\right)^{4/3} = [\overline{(u-u')^2}]^2. \tag{5}$$

p and p' are the pressure contributions at x and $x+r$, respectively. Equation (5) relates the structure function of the longitudinal velocity fluctuation along the flow direction ($\overline{(u-u')^2}$) with the pressure structure function in the same direction ($\overline{(p-p')^2}$).

The coefficient of the velocity structure coefficient can be calculated in two steps from the $-5/3$ power law fit to the measured velocity spectrum. First, the measured longitudinal velocity spectrum in the direction of flow is used to calculate the kinetic energy spectrum $E(k)$. $E(k)$ is the power spectral density of the wind velocity integrated over the entire solid angle in wave number space.

Panofsky and Dutton[6] show that for isotropic homogeneous turbulence the measured velocity one-dimensional power spectral density $F^{11}(k_1)$ is related to the energy spectrum by

$$\frac{d}{dk}\left(\frac{1}{k^3}\frac{dF^{11}(k)}{dk}\right) = \frac{2E(k)}{k^3}. \tag{6}$$

Substitution of the $-5/3$ power law into Eq. (6) gives

$$E(k) = \frac{55}{18}F^{11}(k_1). \tag{7}$$

$F^{11}(k_1)$ is the power spectral density of the wind velocity component in the direction of mean flow measured in terms of the wave number in the direction of flow (k_1). k_1 is calculated from the Fourier frequency of the measured time series by use of Taylor's frozen turbulence hypothesis,

$$k_1 = \frac{2\pi f}{U}, \tag{8}$$

where U is the mean velocity during a measurement set.

Next, the coefficient of the power law in $E(k)$ is related to the coefficient of the structure function. Monin and Yaglom[18] developed and presented the relations necessary to convert power laws in separation distance in the structure function to energy spectrum power laws in wave number magnitude.

Equation 13.100 of Ref. 18 relates the longitudinal structure function of an isotropic vector field of the form,

$$\overline{(u-u')^2} = A_1 r^{2/3}, \tag{9}$$

to an energy spectrum of the form,

$$E(k) = C_1 k^{-(5/3)}, \tag{10}$$

where

$$A_1 = \frac{2\pi C}{\left(3+\frac{2}{3}\right)\Gamma\left(2+\frac{2}{3}\right)\sin\left(\frac{\pi}{3}\right)} = 1.313C_1. \tag{11}$$

A_1 and C_1 are constant coefficients. Equations (7)–(11) are then combined yielding

$$C(\nu\epsilon)^{1/2}\eta^{-(2/3)} = A_1 = 1.313\frac{55}{18}\left(\frac{F_{11}(k_1^0)}{(k_1^0)^{-(5/3)}}\right). \tag{12}$$

The magnitude or height of the power law curve is determined by reading the magnitude of $F^{11}(k_1)$ at a chosen reference wave number k_1^0. The square of Eq. (12) determines the coefficient of the longitudinal pressure structure function.

Similar relations to those above can then be used to calculate the predicted pressure power spectrum in the direction of flow from the pressure power function. Equation 13.59 of Ref. 18 relates the longitudinal structure function of a locally isotropic scalar field, which we will take in the form of Eq. (5),

$$\frac{1}{\rho^2}\overline{(p-p')^2} = A_1^2 r^{4/3}, \tag{13}$$

to the pressure spectrum in terms of wave number magnitude,

$$P(k) = \rho^2 C_1' k^{-(7/3)}, \tag{14}$$

where

$$C' = \frac{\Gamma\left(2+\frac{4}{3}\right)}{\pi}\sin\left(\frac{2\pi}{3}\right)A_1^2 = 0.776A_1^2. \tag{15}$$

The pressure spectral density can then be related to the predicted longitudinal pressure spectra assuming isotropic fields. Lumley and Panofsky[19] developed an equation for relating a scalar one-dimensional pressure spectrum $F^P(k)$ to the corresponding pressure spectrum $P(k)$,

$$\frac{dF^P(k)}{dk} = -\frac{P(k)}{2k}. \tag{16}$$

For $P(k) \propto k^{-(7/3)}$ this gives

$$F^P(k_1) = \tfrac{6}{14}P(k_1). \tag{17}$$

A factor of 2 has been inserted into Eq. (17) relative to Ref. 19 since we are using one-sided spectral densities (i.e., only positive frequencies). The following equation relating the measured one-dimensional power spectral density of velocity along the flow to the predicted power spectral density of pressure is developed by combining Eqs. (5), (7), (12), (15), and (17),

$$F^P(k_1) = \rho^2 \frac{6}{14}\left(\frac{55}{18}\right)^2 (1.32)\left(\frac{F^{11}(k_1^0)}{(k_1^0)^{-(5/3)}}\right)^2 k_1^{-(7/3)}, \tag{18}$$

$$F^P(k_1) = 7.60\left(\frac{F^{11}(k_1^0)}{(k_1^0)^{-(5/3)}}\right)^2 k_1^{-(7/3)}. \tag{19}$$

The factor in parentheses in Eq. (19) is evaluated from a $-5/3$ power law fit to the measured velocity power spectral density as previously described. The numerical factors agree with similar terms in George *et al.*[11] A nominal value of $\rho = 1.2\ \mathrm{kg/m^3}$ has been used in Eq. (19) and throughout this paper as the ambient density of air.

B. Turbulent pressure fluctuation spectrum developed from a LES

Miles, Wyngaard, and Otte[13] developed a computation of turbulent pressure fluctuation spectra based on a LES of turbulence in the atmosphere. The LES is run to produce statistics of the velocity and the turbulence-turbulence pressure fluctuations. The conversion of one-dimensional to energy spectra assumes horizontal isotropy appropriate to atmospheric studies. For forced convection conditions corresponding to turbulent flow they find,

$$F^{P'}(k_1) = 2a_1\epsilon^{(4/3)}k_1^{-(7/3)}, \tag{20}$$

where $2a_1 = 1.68$ with the value determined from their LES. The factor of 2 in Eq. (20) reflects the use of one-sided pressure spectral density such that

$$\int_0^\infty F^{P'}(k_1)dk_1 = \overline{p^2}. \tag{21}$$

The one-dimensional velocity power spectral density is given by

$$F^{11}(k_1) = \frac{18}{55}\alpha\epsilon^{(2/3)}k_1^{-(5/3)}. \tag{22}$$

α is Kolmogorov's constant taken as 1.5, and 18/55 is the factor to relate the three-dimensional energy spectrum to the longitudinal spectrum of the turbulent velocity along the direction of flow. These equations result in a pressure spectral density prediction in terms of the $-5/3$ law fit to the measured velocity spectrum,

$$F^{P'}(k_1) = 10.2\left(\frac{F^{11}(k_1^0)}{(k_1^0)^{-(5/3)}}\right)^2 k^{-(7/3)}. \tag{23}$$

Equations (19) and (23) represent two estimations of the turbulence-turbulence pressure fluctuation in the atmosphere in the inertial range.

III. STAGNATION PRESSURE SPECTRA

Two approaches for calculating the expected power spectral density of the stagnation pressure from the power spectral densities of velocity measurements are presented. The first method is a straightforward one-dimensional application of Eq. (1) with the assumption that the equation applies to time varying velocities. The second is a three-dimensional calculation that closely follows Batchelor's method of calculating the turbulent pressure fluctuation described in Sec. II A.

A. Simple one-dimensional model

In this model, the fluctuating stagnation pressure on a rigid sphere[4,5] is calculated assuming that directional effects can be ignored and that a calculation employing only the fluctuating magnitude will be sufficiently accurate to establish limits on wind noise spectra.

Stagnation pressure is the pressure measured at the zero velocity position on the sphere and is independent of the sphere radius. The time dependent stagnation pressure term is given by Eq. (1),

$$p(t) = \tfrac{1}{2}\rho V(t)^2. \tag{24}$$

We use a Reynold's decomposition to express this in terms of the average flow velocity, U, and the fluctuating velocity magnitude, $u(t)$,

$$p(t) = \tfrac{1}{2}\rho(U + u(t))^2 = \rho U^2 + \rho U u(t) + \tfrac{1}{2}\rho u(t)^2. \tag{25}$$

For outdoor measurements at moderate to high wind speeds, the root-mean-square fluctuation velocity is usually much smaller than the average velocity and we approximate that

$$p(t) \cong \rho U u(t). \tag{26}$$

Converting to the spectral representation gives

$$G^P(k_1) = \rho^2 U^2 F_v(k_1), \tag{27}$$

where $G^P(k_1)$ is the power spectral density of the stagnation pressure and F_v is the power spectral density of the wind velocity magnitude. The measured spectrum is the frequency spectrum calculated from the resultant of a three-component wind velocity measurement sampled in time. This is converted to a wave number spectra using Taylor's frozen turbulence hypothesis and the measured average wind speed, U,

$$F_v(k_1) = \frac{U}{2\pi}F_v'(f), \tag{28}$$

where $k_1 = 2\pi f/U$, and $F_v'(f)$ is the power spectral density of the velocity magnitude in $\mathrm{m^2/s}$.

B. Three-dimensional stagnation pressure calculation

Again we start with the stagnation pressure term but set

$$\mathbf{V} = \hat{i}(U + u_1(t)) + \hat{j}u_2(t) + \hat{k}u_3(t), \tag{29}$$

where U is the average wind speed, $u_1(t)$ is the turbulent velocity in the flow direction, and $u_2(t)$ and $u_3(t)$ are the transverse and vertical turbulent velocity components, respectively. Following Batchelor,[14] the flow is assumed to be

isotropic and homogeneous. Also, the third and fourth moments of velocity are assumed to be related to the second moments as if the distribution were normal. Applying the wind velocity form with the substitution $V^2=\mathbf{V}\cdot\mathbf{V}$ to Eq. (1) gives

$$p(t)=\rho U u_1+\tfrac{1}{2}\rho u_i u_i, \tag{30}$$

where the repeated indices indicate the summation $i=1,3$. This form can be used for stagnation pressure due to cross flows on streamlined microphones by omitting the u_1 terms. Following Ref. 14,

$$\frac{4\overline{pp'}}{\rho^2}=4U^2\overline{u_1u_1'}+2U\overline{u_1'u_iu_i}+2U\overline{u_1u_j'u_j'}+\overline{u_iu_iu_j'u_j'}, \tag{31}$$

where the unprimed quantities are evaluated at the origin and the primed quantities at positive $\mathbf{r}$. The normal moments hypothesis eliminates the third moment terms and the fourth moment terms become

$$\overline{u_iu_iu_j'u_j'}=R_{ii}(0)R_{jj}(0)+2R_{ij}(\mathbf{r})R_{ij}(\mathbf{r}), \tag{32}$$

where $R_{ij}(\mathbf{r})$ is the spatial correlation function of the u_i and u_j components for position separation $\mathbf{r}$. Batchelor[14] shows that for isotropic, homogeneous turbulence with the 1 direction as the flow direction,

$$R_{11}(\mathbf{r})=\overline{u^2}f(r), \tag{33}$$

$$R_{22}(\mathbf{r})=R_{33}(\mathbf{r})=\overline{u^2}f(r)+\tfrac{1}{2}rf'(r), \tag{34}$$

and all off-diagonal correlations are zero. George *et al.*[11] take

$$\overline{u^2}=\tfrac{1}{3}(\overline{u_1^2}+\overline{u_2^2}+\overline{u_3^2}) \tag{35}$$

for flows that may not be completely isotropic. If the turbulence is truly isotropic, $\overline{u^2}=\overline{u_i^2}$ for any i, and Eq. (31) then becomes

$$\frac{4\overline{pp'}}{\rho^2}=4U^2\overline{u^2}f(r)+(\overline{u^2})^2[9+6f(r)+4rf'(r)f(r)+r^2f'^2(r)]. \tag{36}$$

The pressure structure function is formed by noting

$$\overline{(p-p')^2}=2(\overline{p^2}-\overline{pp'}). \tag{37}$$

$\overline{p^2}$ is evaluated by letting $r=0$ in Eq. (36) giving

$$\frac{4\overline{(p-p')^2}}{\rho^2}=4U^2\overline{u^2}(1-f(r))+(2\overline{u^2})[6u(1-f^2(r))-4rf'(r)f(r)-r^2f'^2(r)]. \tag{38}$$

The velocity structure function is given by Eq. (4).

Solving Eq. (4) for $f'(r)$ and substituting in Eq. (38) gives

$$\frac{\overline{(p-p')^2}}{\rho^2}=C(\nu\epsilon)^{(1/2)}\left(\frac{r}{\eta}\right)^{2/3}-\frac{10}{9}C^2(\nu\epsilon)\left(\frac{r}{\eta}\right)^{(4/3)}. \tag{39}$$

This is the one-dimensional structure function for the scalar stagnation pressure. The second term represents a decorrelation due to second order terms in velocity and indicates a $k^{-(7/3)}$ term in the wave number spectrum. The factor $C(\nu\epsilon)^{1/2}\eta^{-2/3}$ is determined from the fit to the measured velocity spectrum by use of Eq. (12).

Monin and Yaglom[18] provide formulas for converting the one-dimensional structure function to the one-dimensional pressure spectrum. The relationships in Ref. 18 are expressed in terms of time and angular frequency, but can be converted to relationships in space and wave number using Taylor's hypothesis. If

$$\frac{1}{\rho^2}\overline{(p-p')^2}=A_3r^\gamma,\quad 0<\gamma<2. \tag{40}$$

then

$$G^{P'}(k)=C_3k^{-(1+\gamma)}; \tag{41}$$

where

$$C_3=\frac{\Gamma(5/3)\sin(\pi/3)}{\pi}A_3\quad\text{for }\gamma=2/3, \tag{42}$$

and

$$C_3=\frac{\Gamma(7/3)\sin(2\pi/3)}{\pi}A_3\quad\text{for }\gamma=4/3. \tag{43}$$

A_3 represents the coefficient of the power law term in Eq. (39). Since $G^{P'}(k)$ is a onedimensional scalar in the direction of flow, it is the measured power spectral density. Gathering terms again,

$$G^{P'}(k_1)=1.14\rho^2\left(U^2+\frac{11}{3}\overline{u^2}\right)\left(\frac{F_{11}(k_1^0)}{(k_1^0)^{-(5/3)}}\right)k_1^{-(5/3)}-5.89\left(\frac{F^{11}(k_1^0)}{(k_1^0)^{-(5/3)}}\right)^2k^{-(7/3)}. \tag{44}$$

This is a second estimate of the stagnation pressure. The relative magnitude of the two contributions depends on the strength of the velocity turbulence in the inertial range and on the wave number. We will evaluate these terms for two measurement examples and explore relative values in Sec. V. Also, note that if the second term is dropped and $\overline{u^2}<U^2$, this equation only differs from Eq. (27) by a factor of 1.14. The velocity spectra are different however; Eq. (27) is calculated from the velocity magnitude spectrum, while Eq. (44) is calculated from the one-dimensional velocity spectra in the flow direction.

IV. SELF-NOISE CONTRIBUTIONS TO THE WIND NOISE SPECTRUM

If the incoming flow is substantially nonturbulent, the turbulence produced by the windscreen or microphone can also produce measurable pressure fluctuations at the microphone. This self-noise provides another limit on the wind noise reduction achievable by nonstreamlined microphones in flow. Strasberg[16,17] has shown that wind noise data taken with spherical and cylindrical windscreens in low turbulence flows can be collapsed into a single curve if the spectral levels are plotted versus the nondimensional frequency, fD/U, where f is the frequency, D is the diameter of the

windscreen in meters, and U is the flow velocity. Strasberg's equation can be expressed in terms of wave number power spectral density as

$$H^P(k) = 3.45 \times 10^{-6} U^4 D^{-2.3} k^{-3.3} \tag{45}$$

for values of dimensionless frequency between 0.1 and 5. The lower limit on the data is the frequency response of the microphones used in the experimental studies. At the upper limit the data no longer collapses into a single line.

Morgan[9] and Hedlin and Raspet[20] have shown that the nondimensional scaling suggested by Strasberg is also useful for studies of wind noise reduction in turbulent flows by collapsing wind noise reduction data from different screens in varying flows into a single curve. The wind noise levels measured in winds near the ground are usually larger than the Strasberg curve. Strasberg notes that this is expected for incoming turbulent flows.

The results of Eq. (45) are displayed in Figs. 3(a) and 3(b). These results represent a limit for wind noise as the incoming turbulence intensity becomes small. It is not clear if the self-noise contribution in a turbulent flow can be predicted by Eq. (45). The windscreen obviously modifies the turbulent field but, to our knowledge, no theory yielding pressure predictions exists. Some of the measurements used by Strasberg were taken using microphones on rotating booms. In this case, the microphone passes through its own wake and the incoming flow would not be entirely nonturbulent. At present, these predictions serve as a bound for the wind noise that would be observed in a substantially nonturbulent flow.

V. EXPERIMENT AND ANALYSIS

The experimental data presented in this paper were recorded on August 25, 2004 at approximately 4:00 p.m. by Shields at Clegg Airport in Oxford, Mississippi. These measurements were not designed to support the theoretical work of this paper. Only runs 14 and 15 out of some 30 runs used microphones/sensors mounted at about the same height as the anemometer. In addition, Shields also placed a sensor inside of a 0.90 m ball of fiberglass home insulation that was then placed on the ground.

Wind velocity measurements were taken using a Campbell Scientific CSAT3 three-dimensional sonic anemometer. This was mounted 1.0 m from the ground and was pointed in the approximate direction of the prevailing winds. Pressure measurements were taken using pressure sensors with various windscreens placed on them. Two of these sensors were Brüel and Kjær (B&K) type 4190 1/2 in. microphones and two were piezoceramic unimorphs embedded in a housing constructed by Shields out of polyvinyl chloride (PVC). These housings were made from PVC end caps glued together to form cylinders approximately 0.1 m in length and 0.07 m in diameter. The unimorphs bisect the cylinders which have approximately 25 one millimeter holes drilled in each end. The unimorphs and housings are designated as the Shields sensor. Detailed descriptions of the unimorphs are provided in Ref. 21.

All the sensors mentioned above were connected to a National Instruments PC based AD/DA card that was then controlled with a program written in Labview©. Each data run lasted 120 s and was taken with a sample rate of 500 samples/s. The frequency responses of the B&K microphones drop off below 5 Hz, and the anemometer had an internal sampling rate of 60 Hz and could measure down to dc. The power spectral densities plotted in Fig. 3 reflect these limitations. The frequency responses of the piezoceramic sensors have not been determined at this time; for completeness, data from these devices span the entire spectrum.

The Shields sensors and the microphones were covered with various windscreens. One microphone was bare except for the standard grid covering, and the other had a 0.18 m diameter open pore foam windscreen with a pore density of approximately 30 pores/in. placed on it. One of the Shields sensors was unscreened, and the other was wrapped in fiberglass home insulation to form a ball of approximately 0.90 m in diameter. The two microphones and the bare sensor were mounted approximately 0.90 m from the ground, while the fiberglass wrapped sensor was placed on the ground. We note that some attenuation of wind noise signal may occur in propagation through the foam or the fiberglass. The attenuation in the 0.18 m foam spheres should be small at the low frequencies of this study, but may be significant in the large fiberglass sphere.

The power spectra presented here are calculated with a standard periodogram. In this approach, the mean squared amplitude of the time series data f, defined by

$$\mathrm{RMS}^2 = \frac{1}{N} \sum_{j=0}^{N-1} |f_j|^2, \tag{46}$$

where N is the number of samples, is equal to the area under the power spectrum curve

$$A_{\mathrm{PSD}}(f_k) = \frac{W_b}{N^2} \sum_{k=0}^{N-1} |F_k|^2, \tag{47}$$

where the F_ks are the binned amplitudes that result from the discrete Fourier decomposition of the time series, and W_b is the width of the bins in which those amplitudes fall. Fourier transforms generate output in unitless bins of size 1, and care must be taken when converting the initial binned amplitudes to either frequency or k space so that the equality of Eqs. (46) and (47) will hold. In other words, to convert from the initial bin width to frequency bins, W_b is multiplied by F_n/N, where F_n is the Nyquist frequency. In order to maintain the value of the summation, each element in the sum must then be divided by F_n/N. Similarly, to convert from frequency to k-space, one must multiply and divide by $2\pi/U$, where U is the average wind speed measured by the CSAT3.

The FFT routines used to generate the power spectra required datablock sizes that were an integral power of 2. Therefore the data sets were divided into 14 blocks of 4096 points. The 2696 points left at the end of each dataset were not used. The Fourier transforms of the blocks were calculated then averaged at each frequency in order to obtain a more reliable statistical estimation of the data. A greater

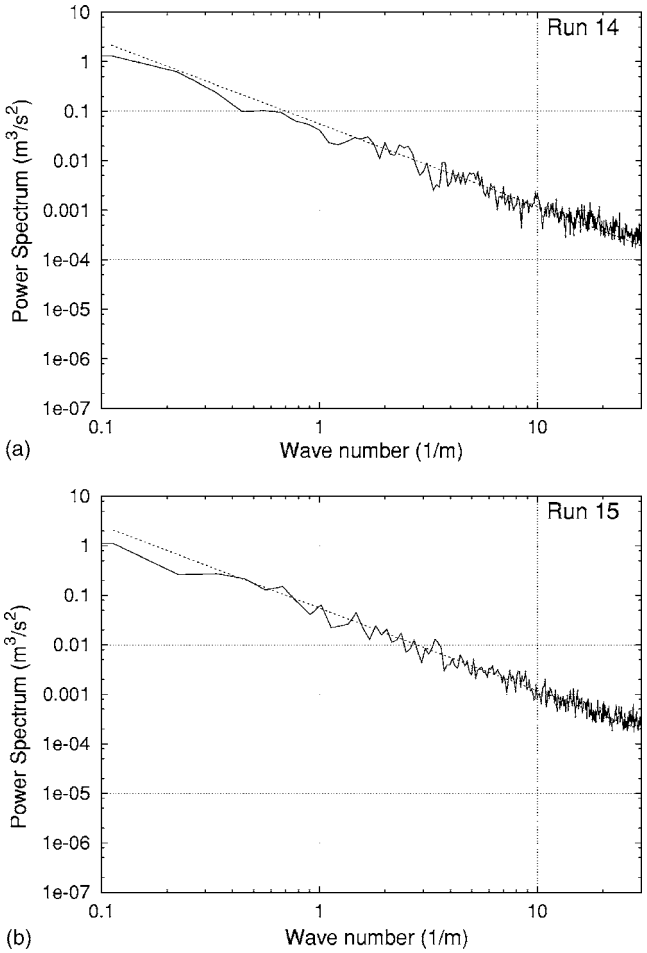


FIG. 1. Wind velocity magnitude power spectral density in wave number along flow direction for (a) run 14 and (b) run 15. Dotted lines are $k^{-(5/3)}$ power law fits to the data in the range $1-30$ m^{-1}.

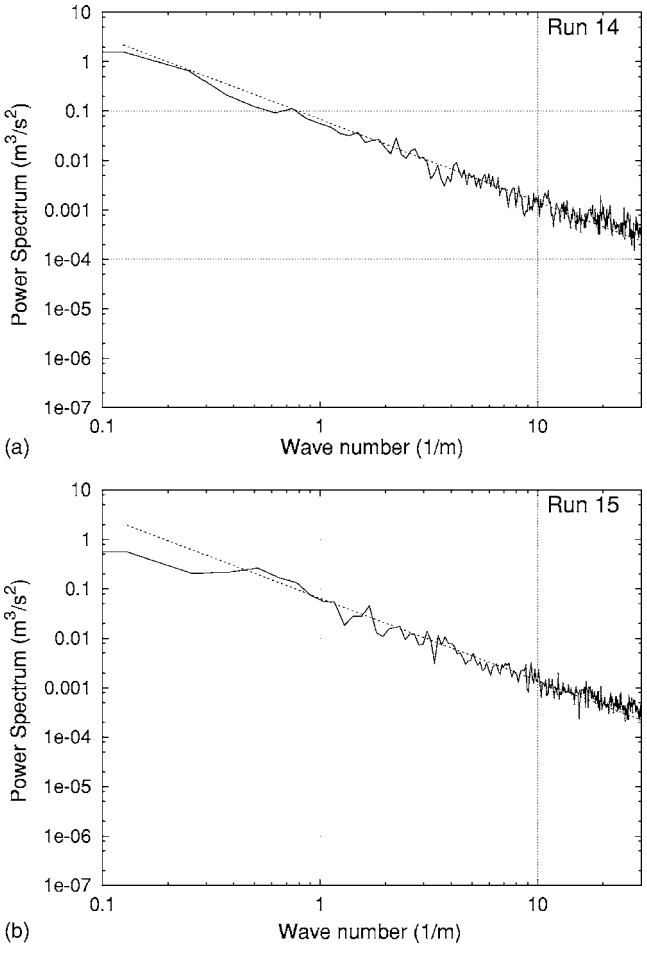


FIG. 2. Wind velocity power spectral density for the wind component in the direction of flow for (a) run 14 and (b) run 15. Dotted lines are $k^{-(5/3)}$ power law fits to the data in the range $1-30$ m^{-1}.

number of averaging over smaller block sizes could have further reduced noise, but would have resulted in less resolution at low frequencies.

The average wind speeds are calculated by taking the mean of the time series data V_i,

$$U = \frac{1}{N}\sum_{i=1}^{N} V_i, \tag{48}$$

where V_i can be either the magnitude of the inflow as defined in Eq. (29), or the component of the velocity data in the direction of flow. Similarly, $\overline{u^2}$ is the square of the RMS of the fluctuating velocity and is calculated by first subtracting the mean flow speed from the time series,

$$\overline{u^2} = \frac{1}{N}\sum^{N} (V_i - U)^2. \tag{49}$$

Values calculated from Eqs. (48) and (49) are used in Eqs. (27), (44), and (45) to generate the stagnation and self noise theory lines in Fig. 3.

Figures 1(a) and 1(b) present the power spectral densities of the wind velocity magnitudes in wave number space. These are calculated by taking the square root of the sum of the squares of the three components from the wind speeds measured with the sonic anemometer. The data are presented up to one-half the sampling frequency of 60 Hz. The fit lines on Fig. 1 are the best fits assuming a $k^{-5/3}$ power law dependence in the wave number range from 1 m^{-1} to 30 m^{-1}. We will assume that the inertial range extends to the upper limit of frequency of the acoustic measurements. Hot wire measurements that are taken at higher frequencies indicate that this is a valid assumption. The fits to the magnitude spectra are used to calculate the predicted one-dimensional stagnation pressure [Eq. (27)].

Figures 2(a) and 2(b) present the power spectral densities of the wind velocity components in the direction of flow, also in wave number space. The data is again well fit by $k^{-5/3}$ within the same wave number range as the magnitude plots. These fits are used in calculating the predicted three-dimensional stagnation pressure [Eq. (44)], the Batchelor prediction of the turbulence-turbulence interaction pressure [Eq. (19)], and the Miles, Wyngaard, and Ott model of the turbulence-turbulence interaction pressure [Eq. (23)].

The values of the fit parameters, average wind speed U, mean square velocity magnitudes, and mean square velocity components are presented in Table I.

Figures 3(a) and 3(b) display the power spectral densities of the pressure fluctuations measured with the sensors and windscreens. The data, in order from highest level to lowest, are for the bare B&K microphone, the Shields sensor,

TABLE I. A listing of the values used in the calculation of the fit lines in Figs. 3(a) and 3(b).

	U(m/s)	$\overline{u^2}$(m^2/s^2)	$F_v(k_1^0)/(k_1^0)^{-(5/3)}$	$F^{11}(k_1^0)/(k_1^0)^{-(5/3)}$
Run 14	6.90	0.59	0.052[a]	0.057[b]
Run 15	6.77	0.47	0.050[c]	0.052[d]
ρ=1.2 kg/m^3				
$D_{\text{windscreen}}$=0.18 m				
$D_{\text{fiberglass}}$=0.90 m				

[a]Determined from the fit line in Fig. 1(a).
[b]Determined from the fit line in Fig. 2(a).
[c]Determined from the fit line in Fig. 1(b).
[d]Determined from the fit line in Fig. 2(b).

the B&K microphone in the 0.18 m windscreen, and the Shields sensor on the ground in a 0.90 m sphere of fiberglass. The predictions of turbulence-turbulence interaction pressure [Eqs. (19) and (23)] and stagnation pressure [Eqs. (27) and (44)] are also displayed in Figs. 3(a) and 3(b). Two

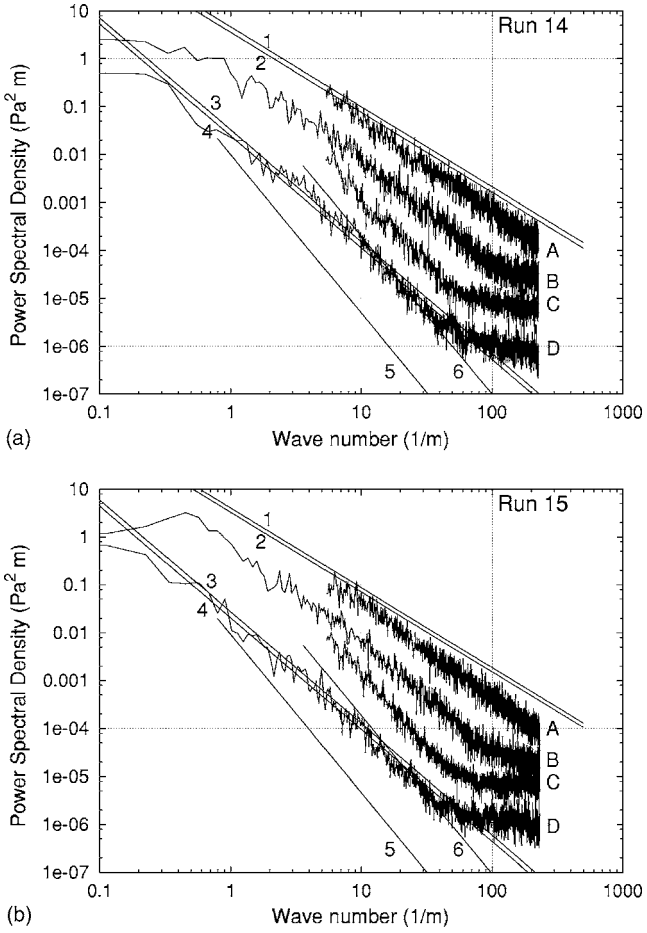


FIG. 3. Prediction lines and power spectral densities of pressure fluctuations in wave number for (a) run 14 and (b) run 15. Data lines are labeled A–D while theory lines are labeled 1–6 and are as follows: Data: A, bare B&K 1/2 in. microphone; B, shields sensor; C, microphone in 0.18 m windscreen; D, Shields sensor in 0.90 m fiberglass windscreen. Predictions: (1) Three-dimensional stagnation pressures [Eq. (44)]; (2) One dimensional stagnation pressures [Eq. (27)]; (3) Batchelor turbulence-turbulence interaction [Eq. (19)]; (4) LES turbulence-turbulence interaction [Eq. (23)]; (5) Self-noise for 0.18 m windscreen [Eq. (45)]; (6) Self-noise for 0.90 m fiberglass windscreen [Eq. (45)].

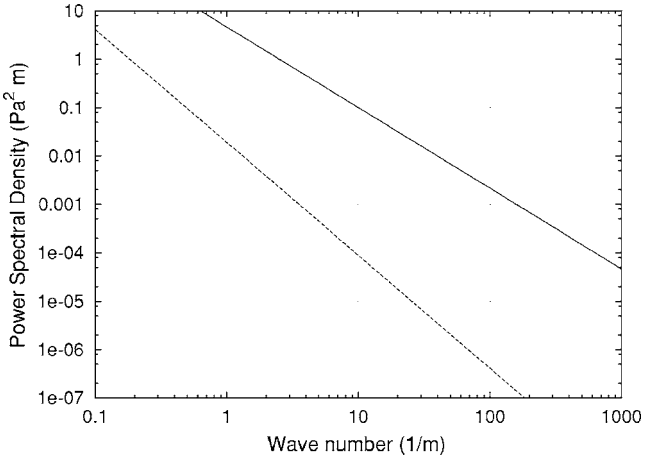


FIG. 4. Magnitude of the $k^{-(5/3)}$ (solid line) and $k^{-(7/3)}$ (dotted line) terms in the three-dimensional stagnation pressure calculation [Eq. (44)].

predictions of self noise [Eq. (45)] are presented: one for the 0.18 m windscreen and one for the 0.90 m fiberglass windscreen.

Figure 4 presents the relative contribution of the first and second terms of Eq. (44). It can be seen that the second term is negligible for our measurements. Spectral levels about 100 times stronger than measured are required for the second term to be comparable to the first.

Figure 5 presents the scaled wind noise reduction of the 0.18 m windscreen referenced to the stagnation pressure as calculated from the velocity magnitude spectrum using Eq. (27). The measured wind noise reduction referenced to a bare microphone measured by Morgan[9] is also presented in this figure. The wind noise reduction of the 0.18 m windscreen agrees reasonably well with Morgan's reduction measurements at low frequencies. At high frequencies, Morgan displays much higher reductions. One cause of the discrepancy is that Morgan's reported bare microphone levels are higher than the stagnation pressure. A second cause may be background acoustic pressure fluctuations in the present study.

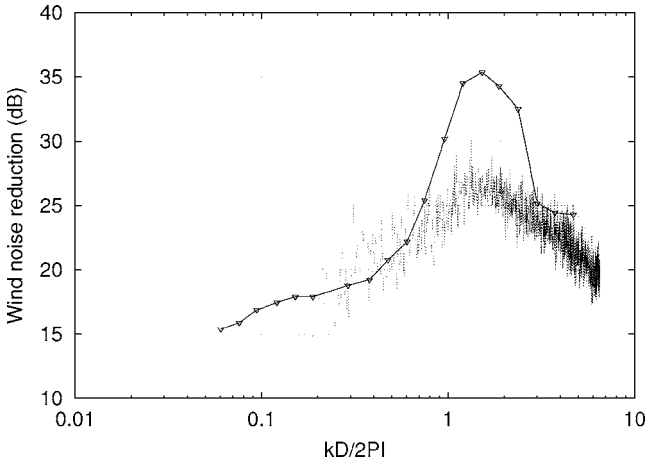


FIG. 5. Plotted versus dimensionless frequency fD/u (cf. Sec. IV) are the wind noise reduction of the 0.18 m windscreen referenced to the fit of the one-dimensional stagnation pressure calculation (dotted line), and the wind noise reduction to a bare microphone from Ref. 8 (connected triangles).

The measured sound levels level off at high frequencies indicating some type of noise floor. These differences will be investigated in future measurement work.

VI. DISCUSSION AND CONCLUSIONS

The principle results of this paper are displayed in Fig. 3. It is encouraging that the two methods of calculating the stagnation pressure agree closely, as do the two methods of calculating the turbulence-turbulence interaction pressure.

The stagnation pressure calculation provides an upper bound on the wind noise measurements made with a bare microphone. The bare microphone data, displayed as curve A in Fig. 3, was taken with a grid in place over the diaphragm. For this measurement, the grid acts like a small windscreen and reduces the wind noise level slightly below the stagnation pressure prediction. The wind noise reduction of the grid increases slightly with increasing wave number.

The two calculations of the turbulence-turbulence interaction pressure lie along the pressure measurement of the large ball (0.90 m) of fiberglass lying on the ground. The pressure predictions are calculated at the height of the anemometer while the Shields sensor at the center of the ball is only at 0.45 m. The agreement of theory and measurement in this case is purely accidental. The large fiberglass screen is not at the same level as the predictions and will present significant attenuation of wind noise pressure fluctuations or acoustic signals. The results do illustrate the level of effort and dimensions of windscreen necessary to reduce the wind noise to the approximate level of the turbulence-turbulence interaction pressure in the inertial range.

It is clear that either stagnation pressure calculation provides a good estimate of the maximum wind noise given the average wind speed and wind velocity spectral level at the same height. With the limited data presented here, it is not possible to determine if Eq. (44) contains more physics than Eq. (27), or is just more complicated. Such a determination requires careful measurements with transducers set into spheres and measurements under different atmospheric conditions with greater variation in U and $\overline{u^2}$.

As such, we propose that the simpler Eq. (27) be used with wind velocity magnitude spectra as the reference maximum wind noise in wind noise reduction studies (see Fig. 5). This choice eliminates uncertainties due to the wind noise of the bare microphone, grid and mounting hardware. This technique also eliminates variations in reduction due to variations in the relative turbulence level due to weather conditions. The wind noise depends on both the wind speed and the turbulence level as determined by Eq. (27) and by Morgan.

The close agreement of the two theories for calculating the turbulence-turbulence interaction spectra and magnitude serve to validate the basis of each theory. The approximations employed are quite different, but the predictions are well within a factor of 1.5. Batchelor's theory[14] has been verified using careful measurements in a controlled turbulent flow by George *et al.*[11]

Physically, the turbulence-turbulence interaction pressure fluctuations can be thought of as moving regions of high and low pressures convected at the wind speed. For a given frequency the size of the pressure region is on the order of U/f, where f is the frequency and U the average wind speed. A compact wind screen made of low resistivity foam or fiberglass will have a negligible effect on the quasistatic intrinsic pressure fluctuations. At higher frequencies where the turbulence size is on the order of the windscreen diameter or smaller, averaging over multiple turbulence regions can occur and the measured fluctuation pressures can be smaller than the predicted turbulence-turbulence interaction pressures. As such, these predictions of the turbulence-turbulence interaction pressure represent an upper limit on wind noise reductions achievable with compact windscreens. The results displayed in Fig. 3 indicate that the maximum reduction achievable for our measurement conditions with a compact screen is about 20 dB at $k=1.0\ \mathrm{m}^{-1}$ and about 35 dB at $k=100\ \mathrm{m}^{-1}$. The maximum reduction is smaller if the relative turbulence level is higher since turbulence-turbulence pressure is proportional to $F^{11}(k_1^0)/(k_1^0)^{-(5/3)}$ squared and the stagnation pressure is proportional to the product of $F^{11}(k_1^0)/(k_1^0)^{-(5/3)}$ and the average velocity U.

We note that Shields[21] describes a number of tests with his sensor in Ref. 21 and concludes that he is measuring the true fluctuating pressure in the atmosphere, unaffected by the sensor's presence. His measurements are a factor of 20 dB above the predictions of the turbulence-turbulence interaction based on the calculations of this paper. We believe that the Shields sensor measures pressure fluctuations induced by the presence of the bluff cylinder section in the fluctuating flow, and not the true fluctuating pressure created by the turbulence-turbulence interaction.

Strasberg's self noise predictions[17] lie below the measured data for the corresponding sensors indicating that self-noise is not a significant contribution for the measurements shown. Since this equation is empirical, the corresponding lines in Fig. 3 terminate at the lowest frequency considered by Strasberg. If the line is extrapolated to lower wave numbers, the self-noise prediction would intersect the turbulence-turbulence prediction at low wave numbers. Under low turbulence conditions, this contribution may not be negligible, particularly at small wave numbers. However, Strasberg argues that the lowest frequency measured should be on the order of a dimensionless frequency of 0.2, implying that extrapolation would not be valid. This question can only be resolved through further experimentation.

In this paper, theories for the turbulence-turbulence interaction pressure in the inertial range have been adapted for the study of wind noise in outdoor microphones. These predictions provide a lower limit for wind noise for sound measurement using compact devices. To achieve lower wind noise levels than the turbulence-turbulence interaction pressure, averaging over spatial regions comparable to the turbulence scale is necessary. Theories for the maximum expected wind noise corresponding to a given velocity fluctuation spectrum have also been developed. The maximum level can serve as the reference level for reporting wind noise reduction. Using this reference would eliminate experimental uncertainties due to differences in geometry and mounting of bare microphones. These two sets of theories span the range

of wind noise spectral data collected with sensors varying from bare microphones to large fiberglass insulation windscreens. These equations and the measurement techniques provide a solid framework for future research in wind noise reduction outdoors. Extension of the theories to wave numbers below the inertial range and to streamlined probes will be pursued.

ACKNOWLEDGMENTS

This work was supported in part through collaborative participation in the Collaborative Technology Alliance for Advanced Sensors sponsored by the U.S. Army Research Laboratory under Cooperative Agreement No. DAAD19-01-0008. We wish to thank Dr. F. D. Shields of Miltec Research and Technologies Incorporated for providing us with the data used in this paper and for many stimulating discussions concerning wind noise, Dr. Carrick Talmadge of the NCPA for providing guidance on the spectral analysis, and Dr. John Wyngaard of Penn State University who provided help in understanding the analysis in his paper. We also wish to recognize the value of discussions with Dr. Keith Wilson of the U.S. Army Cold Regions Research Lab and with Dr. Henry Bass of NCPA.

[1]S. Morgan and R. Raspet, "Investigation of the mechanisms of low-frequency wind noise generation outdoors," J. Acoust. Soc. Am. **92**, 1180–1183 (1992).

[2]R. S. McGuinn, G. C. Laucle, and D. C. Swanson, "Low flow noise microphone for active noise control applications," AIAA J. **35**, 29–34 (1997).

[3]M. R. Shust and J. C. Rogers, "Active removal of wind noise from outdoor microphones using local velocity measurements," J. Acoust. Soc. Am. **104**, 1781 (1998).

[4]O. K. Rediniotis and R. E. Kinser, "Development of a nearly omnidirectional velocity measurement pressure probe," AIAA J. **36**, 1854–1860 (1998).

[5]R. J. Dobosy, T. L. Crawford, D. L. Auble, G. H. Crescenti, and R. C. Johnson, "The extreme turbulence (ET) probe for measuring boundary-layer turbulence during hurricane-force winds," Eleventh Symposium on Meteorological Observations and Instrumentation, Albuquerque, NM, Amer. Meteor. Soc. (2001), pp. 50–54.

[6]H. A. Panofsky and J. A. Dutton, *Atmospheric Turbulence, Models and Methods for Engineering Applications* (Wiley, New York 1984).

[7]W. Phelps, "Microphone wind screening," RCA Rev. **3**, 203–212 (1938).

[8]Z. C. Zheng and B. K. Tan, "Reynolds number effects of flow/acoustic mechanisms in spherical windscreens," J. Acoust. Soc. Am. **113**, 161–166 (2003).

[9]M. S. Morgan, "An investigation of the sources and attenuation of wind noise in measurement microphones," Ph.D. dissertation, University of Mississippi, 1993.

[10]P. D. Schomer, R. Raspet, J. Brunner, D. Marshall, M. Wagner, and D. Walker, "Reduction of wind noise for unattended blast noise monitoring," Noise Control Eng. J. **34**, 77–88 (1990).

[11]W. K. George, P. D. Beuther, and R. E. A. Arndt, "Pressure spectra in turbulent free shear flows," J. Fluid Mech. **148**, 155–191 (1984).

[12]H. V. Fuchs, "Measurement of pressure fluctuations within subsonic turbulent jets," J. Sound Vib. **22**, 361–378 (1972).

[13]N. Miles, J. C. Wyngaard, and M. J. Otte, "Turbulent pressure statistics in the atmospheric boundary layer from Large-Eddy Simulation," Boundary-Layer Meteorol. **113**, 161–185 (2004).

[14]G. K. Batchelor, "Pressure fluctuations in isotropic turbulence," Proc. Cambridge Philos. Soc. **47**, 359–374 (1951).

[15]J. A. Elliot, "Instrumentation for measuring static pressure fluctuations within the atmospheric boundary layer," Boundary-Layer Meteorol. **2**, 476–495 (1972).

[16]M. Strasberg, "Nonacoustic noise interference in measurements of infrasonic ambient noise," J. Acoust. Soc. Am. **66**, 1487–1493 (1979).

[17]M. Strasberg, "Dimensional analysis of windscreen noise," J. Acoust. Soc. Am. **83**, 544–548 (1988).

[18]A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics: Mechanics of Turbulence* (MIT Press, Cambridge, 1975).

[19]J. L. Lumley and H. A. Panofsky, *The Structure of Atmospheric Turbulence* (Interscience, New York, 1964).

[20]M. A. H. Hedlin and R. Raspet, "Infrasonic wind-noise reduction by barriers and spatial filters," J. Acoust. Soc. Am. **114**, 1379–1386 (2003).

[21]F. Douglas Shields, "Low-frequency wind noise correlation in microphone arrays," J. Acoust. Soc. Am. **117**, 3489–3496 (2005).

Acoustic detection and quantification of benthic egg beds of the squid *Loligo opalescens* in Monterey Bay, California

Kenneth G. Foote
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Roger T. Hanlon
Marine Biological Laboratory, Woods Hole, Massachusetts 02543

Pat J. Iampietro and Rikk G. Kvitek
California State University Monterey Bay, Seaside, California 93955

(Received 12 May 2005; revised 13 October 2005; accepted 15 November 2005)

The squid *Loligo opalescens* is a key species in the nearshore pelagic community of California, supporting the most valuable state marine fishery, yet the stock biomass is unknown. In southern Monterey Bay, extensive beds occur on a flat, sandy bottom, water depths 20–60 m, thus sidescan sonar is a *prima-facie* candidate for use in rapid, synoptic, and noninvasive surveying. The present study describes development of an acoustic method to detect, identify, and quantify squid egg beds by means of high-frequency sidescan-sonar imagery. Verification of the method has been undertaken with a video camera carried on a remotely operated vehicle. It has been established that sidescan sonar images can be used to predict the presence or absence of squid egg beds. The lower size limit of detectability of an isolated egg bed is about 0.5 m with a 400-kHz sidescan sonar used with a 50-m range when towed at 3 knots. It is possible to estimate the abundance of eggs in a region of interest by computing the cumulative area covered by the egg beds according to the sidescan sonar image. In a selected quadrat one arc second on each side, the estimated number of eggs was 36.5 million. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149840]

PACS number(s): 43.30.Gv, 43.30.Sf, 43.30.Xm, 43.80.Jz [WMC] Pages: 844–856

I. INTRODUCTION

A. Ecological and economic importance of *Loligo opalescens*

Squids, which are cephalopod mollusks, are key players in many marine ecosystems, both as predator and prey.[1] *Loligo opalescens* is a major prey species in the central and southern California coastal ecosystem. In Monterey Bay, juvenile and adult squids of this species are important food items in the diets of 19 fish, nine bird, and two marine mammal species.[2] According to this enumeration, *Loligo opalescens* has the greatest number of predator species in the particular nearshore pelagic ecosystem.

Loligo opalescens supports the most valuable marine fishery in California. This fishery is 140 years old,[3,4] has quadrupled in the past decade, and is estimated to be at maximum exploitation, yet no fishery management plan has existed until autumn 2004.[5,6] The management plan is based largely on an "egg escapement model,"[7] and thus it might be useful to learn the extent of actual egg laying in any given season or year. Most loliginid squids, such as *Loligo opalescens*[3] and *Loligo vulgaris reynaudii*,[8] are among the very few marine species that deposit their gelatinous egg capsules in mops directly on the substrate, thus rendering their spatial distribution immediately accessible to direct measurement.

Timely, synoptic data on the spatial distribution and abundance of *Loligo opalescens* in Monterey Bay are of particular interest both for management and for understanding the ecology, including that of the larger Monterey Bay National Marine Sanctuary. Given accessibility of the squid, acoustics would appear to be a useful tool for such work.

B. Overview of acoustic methods to quantify squid

Traditionally, several direct and indirect methods have been employed to assess squid stocks. These methods include, for example, those based on catch-per-unit effort (CPUE) and swept area of bottom trawl, with various applications to *Loligo pealeii*,[9] *Loligo forbesi*,[10] and *Loligo gahi*.[11] Known difficulties in assessing demersal fish by bottom trawl[12,13] also apply to squid. Use of pelagic trawl or plankton net is also difficult, as when surveying cephalopod larvae, including *Loligo opalescens* larvae,[14] to achieve adequate areal coverage. Some aspects of fishery dynamics have been measured remotely to yield a proxy for fishing effort, thence enabling abundance to be assessed indirectly through a landings-per-unit-of-effort (LPUE) measure.[15] The method of biomass estimation by enumerating squid beaks in the stomachs of predators is known to be problematical.[16] For *Loligo opalescens*, egg and larval surveys have not succeeded.[17] The potential of camera-based surveys has been recognized, but principally for quantification of benthic eggs rather than swimming animals that may avoid lights.[17] Other, generic difficulties apply, including that of efficiency when having to survey a rather large area with an instrument dependent on underwater visibility. Advantages of the direct acoustic surveying method for general fisheries applications are thus appreciated: they are rapid, remote, noninvasive, synoptic, and quantifiable.[13,18] General disadvantages of

acoustic methods are associated with detection, identification, and knowing the applicable characteristic scattering properties of individual organisms in order to be able to reduce acoustic measures of organism density to biological measures of concentration density. Specific disadvantages for squid according to Thorne and Starr[19] are those of near-bottom detection, identification without having recourse to other means, and quantification in the presence of avoidance behavior or in the case of dense schools causing extinction.

The feasibility of estimating squid stock abundance by acoustic means has been established in a number of studies, including demonstrations through sea trials. Examples are *Loligo gahi* and *Martialia hyadesi* on the Falkland Shelf in the South Atlantic[20] and *Loligo vulgaris reynaudii* off the south coast of South Africa.[21] Acoustic data were derived from echo sounders with vertically oriented beams at 38 and 120 kHz for the first two species and at 38 kHz for the third species. In recent years, the abundance of *Todarodes pacificus* on the major fishing ground off northern Honshu, Japan, has been estimated at 38 kHz by the echo integration method (Ref. 22 and A. Kawabata, pers. comm.). Acoustic trials are also being performed on the same species around Hokkaido, Japan, and on *Loligo edulis* in the East China Sea (A. Kawabata pers. comm.).

The acoustic detectability of *Loligo opalescens* aggregations in the water column was established by Vaughan and Recksiek[23,24] in Monterey Bay. Several distinctive echo types were observed at 38, 50, 75, and 200 kHz. These were associated with proximity to the bottom and time of day.

Loligo opalescens has also been observed acoustically off the central Oregon coast by Jefferts *et al.*[25] This was quantified by means of integration of echo sounder data at 120 kHz. In addition, the target strengths of resolved individual squid were measured directly by means of the dual-beam function of the same echo sounder.[26]

C. Biology of *Loligo opalescens*

Loligo opalescens is very fast-growing, short-lived, and highly mobile, as described in, for example, Refs. 3 and 27–30. Little is known in detail about its whereabouts for much of its life history. The exceptions are the end points defined by spawning. Near the end of its 6–12 month life span, *Loligo opalescens* migrates inshore and spawns, laying eggs in capsules on the shallow coastal seafloor. After about one month, paralarvae, which are morphologically similar to the adult, emerge from the capsules to enter the plankton stage; their movements are generally unknown.[31] Those individuals that survive predation and other forms of mortality return as adults to spawn. Significantly for acoustic investigations, the egg laying seems to occur mainly on flat sandy bottoms in depths of 20–60 m.[31–37]

The eggs are laid in gelatinous capsules containing 100–200 eggs, each in its own Voronoi-like cell[38] called a chorion. Initially, the capsules are cylindrical in form, *ca.* 16 mm in diameter and about 9–10 cm long, expanding to 20–25 cm after a few days.[39] The female cements each capsule in the sand, and they stand vertically, bending and swaying under the influence of currents. The presence of egg capsules attracts other spawning squid, which tend to deposit their egg capsules contiguously with existing capsules. In this way, benthic *egg mops* of diameter 20–100 cm and *egg beds* of diameter greater than 1 m are formed (Fig. 1). For simplicity, these are often referred to generically as egg beds. These can become massive, reaching 12 m in extent,[39] and may even consist of multiple tiers rising 40–60 cm off the sea floor or, as reported anecdotally in Ref. 3, in a pillar 1.5–m high. Hatching occurs about one month after egg-laying, depending upon temperature.[28] The period of peak spawning in Monterey Bay is thought to coincide with peak fishery catches during April–May. Squids continue to be fished through the summer, with a secondary capture peak in September–October.[6] Eggs are present throughout this period. One record exists of the presence and absence of eggs following an El Nino event;[40] however, no maps are published of egg distribution or abundance, either yearly or seasonal.

The primary spawning grounds of *Loligo opalescens* in central California are located in the southern part of Monterey Bay over an area that is about 10 km^2. This is small compared to the larger habitat, but is still large for detailed surveying. Currently, the State of California closes

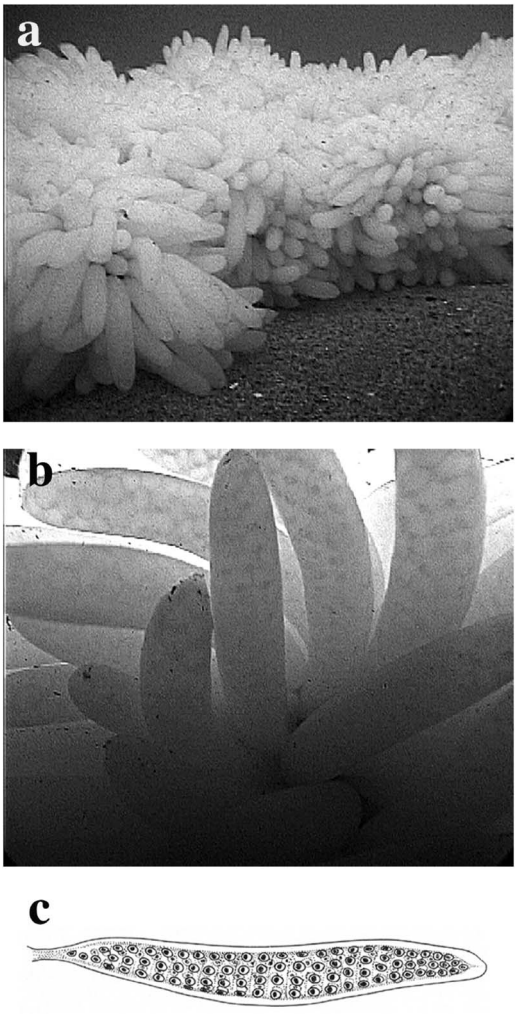


FIG. 1. Egg bed of the squid *Loligo opalescens*. (a) Section of egg bed. (b) Detailed view. Individual chorions can be seen. (c) Sketch of egg capsule [redrawn from C. F. E. Roper, Bull. Mar. Sci. **15**, 589–598 (1965)].

squid fishing two days per week,[35] but no benthic habitat for squid egg beds is protected. Successful annual recruitment is required to sustain squid populations.[41] If fishing pressure continues to increase, it may be prudent to restrict fishing directly over the largest egg beds to allow the normal progression of sexual selection and egg-laying behaviors.[34,37]

D. Acoustic technology

The unique loliginid habit of laying massive benthic egg beds can be exploited in surveying egg beds and thus using egg capsule production as a proxy for reproductive success. Certainly the advantages of surveying a stationary resource and performing analyses based on spatial information are both prominent and tempting in choosing a method. To meet the requirements of resolution and detectable backscatter, it would seem that a sonar with a very directional beam and rather low grazing angle relative to the bottom would be most effective; this defines a towed, high-frequency sidescan sonar.[42,43]

E. Outline of this paper

The plan of this work is to describe development of an acoustic method to detect, identify, and quantify egg beds of *Loligo opalescens*. In the following, preliminary investigations on squid egg capsules and beds are summarized. The design of an experiment to verify the acoustic method is elaborated. Details are given on the instrumentation, equipment, platforms, and methods. Video and acoustic images of the same regions of interest are included in the results. The abundance of eggs is estimated in a quadrat that is one arcsecond on each side. Outstanding problems connected with acoustic resolution, use of alternate platforms, automated recognition of egg-bed echoes, squid biology, monitoring of squid spawning grounds, and acoustic properties of egg capsules are discussed.

II. PRELIMINARY ACOUSTIC OBSERVATIONS OF SQUID EGG BEDS

Acoustic detectability of egg capsules of *Loligo pealeii* was established at the Marine Resources Center, Marine Biological Laboratory, by two of the authors (KF and RH), together with J. Condiotty and C. Gubala in June 2002. Egg capsules freshly laid by the squid in a mop on the bottom of a tank, water depth 80 cm, were observed with the Simrad EY500/710-kHz echo sounder, with transducer floating on the water surface, using a short-duration pulse. *Loligo pealeii*, like *Loligo opalescens*, attaches its egg capsules to the bottom substrate, preferring to abut previously laid capsules. Relative to the echo from the bottom of the tank, the echo from the capsules arrived earlier by a time corresponding to the height of the capsules, 8 cm. Despite the particular *ex situ* conditions, this measurement gave tangible proof that egg capsules of squid could be detected acoustically.

In a study by Sauer *et al.*,[44] some evidence was presented for the acoustic identification of *Loligo vulgaris reynaudii* egg beds *in situ*. A commercial fisheries echo sounder, with operating frequency of 200 kHz, was used. While the identification probability was very high under the circumstances, it is not clear that the significant part of the echo was due to the egg beds themselves and not to bottom morphological features or spawning squid in the vicinity of the egg beds.

The initial acoustic investigation in the field was conducted in May 2003 in Monterey Bay. A 400-kHz sidescan sonar was towed from the California Department of Fish and Game (CDFG) research vessel Mako. Within minutes of commencing the first sidescan sonar transect off Del Monte Beach, with bottom depth of 20–30 m, a distinct dapple or mottling was observed on the sidescan sonar echo record. The size, shape, and proximity of the small features were reminiscent of squid egg beds that SCUBA divers in the scientific party had earlier observed.

Because of the proximity of the beach and exposure of the site to the open ocean, a second site was visited, just north of the breakwater and several hundred meters off Cannery Row. The sidescan sonar observations were similar to those off the beach. SCUBA divers descended to confirm a large concentration of squid egg beds. Various attempts were made to establish a close correlation of the sidescan sonar recordings and squid egg beds. Navigation was a limiting factor, but the results were strongly suggestive.

Efforts to detect the squid egg beds with the Simrad EK60 scientific echo sounder,[45] the successor to the EK500 scientific echo sounder,[46] with split-beam transducer at 200 kHz, were unsuccessful. Echoes from the egg beds and bottom could not be distinguished, which is not surprising, given the pulse length 10 cm, nominal 9–10 cm height of newly laid egg capsules, bottom depth of 20–30 m, sea state, and vessel mounting of the transducer.

A second acoustic investigation in the field was performed about the Channel Islands in southern California in late February and early March 2004. A 600-kHz sidescan sonar was towed from the NOAA Channel Islands National Marine Sanctuary research vessel Shearwater. This was configured with a video camera, with downward orientation. Squid egg beds were observed simultaneously with both systems, if at different ranges. An unexpectedly strong current dragged a bottom-mounted array of acoustic reflectors including floats, out of alignment, preventing repeated observations of the reflectors and squid egg beds.

III. EXPERIMENTAL DESIGN

Based on the preceding investigations, with their suggestive findings and association of certain characteristic features on the sidescan sonar image and squid egg beds, it was decided to perform a larger-scale verification trial on the spawning grounds in Monterey Bay. The principal acoustic instrument would be a high-frequency sidescan sonar with integrated geo-referencing processing system. The major verification tools would be video cameras. One would be a drop camera, to be suspended over the side of the vessel for reconnaissance use, and the second would be ROV-borne for investigations with a relatively high degree of operator control.

Initially, a region of high concentration of egg beds had to be found. This was done by means of the drop camera

over traditional spawning areas. A survey region of interest (roi) was defined. This was surveyed systematically by sidescan sonar along parallel transects, as in Refs. 47 and 48, with substantial overlapping coverage to enable a map, or mosaic, of seafloor backscattering to be assembled.

Locations with markings suggestive of squid egg beds were noted. Following the sidescan sonar survey, the same roi was surveyed by means of a ROV-borne video camera. The degree of coverage was much less, but the entire roi was sampled along a few transects. The ROV was steered to investigate previously identified areas of special interest.

Precision navigation instrumentation was used. A differential Global Positioning System (dGPS) was used to determine the positions of the sidescan sonar transducer arrays and ROV-borne video camera. An acoustic tracking system was used for precise determination of the position of the video camera relative to the vessel.

Given both sonar and video data on the roi, tests were devised to prove that characteristic sidescan sonar features could be associated unambiguously with the presence or absence of egg beds. Two tests were designed, one with the sidescan sonar data enabling predictions to be made for verification by the video-camera images, and *vice versa*. Statistics could then be employed to assign significance levels.

In addition, the sidescan sonar data provided the grist for quantification of the egg beds within a defined area. This served as an example of the feasibility of abundance estimation.

IV. MATERIALS

A. Acoustic instrumentation

1. Sidescan sonar

The principal acoustic instrument was the EdgeTech model 260-TH Image Correcting Side Scan Sonar, with model 272-TD dual-frequency analog towed vehicle, called a towfish, with time-varied gain (TVG). This was used to detect and image the squid egg beds. Two longitudinally oriented, parallel transducer arrays were mounted on each side of the towfish, with nominal operating frequencies of 100 and 500 kHz, as casually described by the manufacturer, but with the higher frequency being much closer to 400 kHz. A deck unit supplied power to the towfish, controlled this, printed the paper record, and sent the analog signal to the Isis Sonar digital acquisition system contained in the model 260-TH Image Correcting Side Scan Sonar.

Both frequencies were used. The upper frequency was in the range 390±20 kHz. The pulse duration was 0.01 ms. At 50-m range, the ping interval was 75 ms, hence the ping rate was 13/s.

The sampling frequency was 2048 16-bit pixels per channel per second, as determined by the AU-32 A/D board in the Isis acquisition system, not the sonar, which was entirely analog. The source level was 222 dB *re* 1 μPa at 1 m. The horizontal beamwidth was 0.5 deg at the −3 dB level, and the vertical beamwidth was 50-deg, tilted downward by 20 deg from the horizontal.

The lower frequency was 105±10 kHz. The pulse duration was 0.1 ms. At 50-m range, the ping interval, ping rate, and sampling frequency were the same as for the higher frequency, but the source level was 228 dB *re* 1 μPa at 1 m. The horizontal beamwidth was 1.2 deg, and the vertical beamwidth was 50 deg, tilted down 20 deg from the horizontal.

2. Fishing echo sounder

A Furuno Fish Finder, model FCV582L echo sounder, was used for general inspection of the water column and occasional detection of squid aggregations in advance of sonar or video observations of egg beds and part of the water column. The echo sounder operated at both 50 and 200 kHz, with a nominal maximum range of 750 m at the lower frequency, with 600-W output power. Echo data were converted to a video signal and displayed on a color liquid crystal display (LCD). Automatic gain control was used but not monitored.

B. Video cameras

1. Drop camera

The reconnaissance camera was suspended on a weighted line from the vessel. This was the Deep Blue Pro Color camera, which forms a National Television System Committee (NTSC) composite video image with 480 TV lines. This was a compact unit, with 3.6-mm wide-angle lens, focus fixed at 2.54 cm to infinity, sensitivity 0.3 lux, and electronic iris. Image data were transferred by cable to the vessel and stored on a JVC BR-DV600A mini-digital video camera recorder (VCR). Telemetry and other information were overlaid on the video image.

2. ROV-carried camera

The camera used in the verification exercises was that mounted inside the remotely operated vehicle (ROV): a JVC model with 470-lines resolution, 0.95–lux color charge-coupled device (CCD) and F 0.8 Pentax lens. Through its internal mounting, it could tilt and view through 360 deg. Panning of the camera was achieved by rotating the vehicle, which could turn on its own axis. Image data were transferred through the ROV tether and, as with the drop camera, recorded on an on-board JVC BR-DV600A mini-digital VCR.

C. Platforms and positioning devices

1. Research vessel MacGinitie

The principal platform was R/V MacGinitie, an aluminum hydrographic survey launch, overall length 9.6 m, draft 0.5 m, beam 2.7 m, gross weight 4.5 metric tons fully equipped, cruising speed 18–28 knots depending on conditions, electrical power 30-A 110-V ac and 12-V dc. It carried state-of-the-art sonar mapping and ROV technology like that found on large hydrographic research vessels, but in a towable configuration designed for nearshore shallow-water habitat work. It is owned and operated by the Seafloor Mapping Laboratory of California State University, Monterey Bay.

2. Remotely operated vehicle Hyball

A platform for the JVC video camera was the Hyball ROV, which has a 300-m depth capability and is fitted with four thrusters. Two main thrusters provide forward, reverse, and rotational movement, and two vertran thrusters provide vertical and lateral movement. This small unit was easy to deploy and retrieve and proved sufficiently maneuverable in depths up to 34 m with mild or no current.

D. Positioning systems

1. Global Positioning System

Research vessel MacGinitie has a differential GPS (dGPS). This is fully integrated with the PC-based Nobeltec/Sitex navigation system of the vessel, which also includes digital charting, radar, and autopilot. Data from the dGPS could be attached to the sonar and video data at the time of their collection.

2. Trackpoint

The Hyball ROV was tracked and its position determined by a Trackpoint II+ ultra-short baseline acoustic tracking system (ORE International, Houston, Texas). The Trackpoint system has a 0.5-m slant-range accuracy and a total horizontal position accuracy of less than 0.5% root-mean square of slant range. Depth information was provided by a pressure sensor on the vehicle as well as by a 4337B Trackpoint multibeacon with depth telemetry. The Trackpoint system was interfaced to the top-side vessel survey navigation system with dGPS and provided real-world (x,y,z)-positioning of the ROV with accuracy ±2 m by means of the Coastal Oceanographics Hypack Max software. The National Marine Electronics Association (NMEA) position data were also recorded directly on the digital video (DV) tape using a Horita GPS-3 encoder. This device recorded the position data in an audio track on the videotape, where they could be accessed without obscuring the video imagery.

V. FIELD METHODS

The field work was conducted on the traditional spawning grounds of *Loligo opalescens* in southern Monterey Bay during the period 13–16 May 2004. This work began with a reconnaissance with the drop camera to locate a region abundant in egg beds. The identified region was then surveyed systematically by sidescan sonar. It was subsequently surveyed by ROV-borne camera to verify the sidescan sonar observations.

A. Drop-camera reconnaissance

The initial drop-camera reconnaissance was performed on 13 May over a flat sandy bottom off the breakwater and in view of Cannery Row at four stations with the vessel drifting at 0.5–1 knot. The approximate depth range was 26–34 m. Small numbers of lone squids were observed scattered widely on the bottom, generally in the vicinity of eggs, but neither adults nor eggs were numerous.

The reconnaissance was continued on 14 May, drifting from north of Lovers Point to Hopkins Marine Station at the approximate speed 0.5–1 knot. The drift path followed that of a fault line, depth difference 6–7 m, dividing an otherwise flat sandy bottom within the depth range 25–40 m. The drift path was maintained near the fault line by use of the vessel motor when necessary. Sparsely and richly populated patches of squid egg beds were observed.

On both days notable features were marked. This operation generated a computer file with paired geographical positions and time codes, as well as biological annotations. The drop-camera position was assumed to be identical to that of the vessel since the camera was suspended from the drifting vessel over a rather short distance, and the line was observed to be vertical.

B. Sidescan sonar survey

The fault-line area identified by the drop-camera reconnaissance was surveyed with the sidescan sonar on 15 May (Fig. 2). Twenty parallel transects were performed at the speed 2–4 knots with the maximum range of 50 m to each side of the towed arrays. The sidescan sonar was maintained about 5–10 m over the seafloor, whose depth varied over the approximate range 25–40 m. The first 15 transects were performed at the upper frequency, about 400 kHz, and the last five at the lower frequency, about 100 kHz. The first 15 transects were performed with a track spacing of 50 m, ensuring 100% overlap in the port or starboard swath. Working conditions were characterized by a fresh onshore breeze and short, steep seas of wave amplitude 1–2 m.

C. ROV-borne video camera investigations

At the beginning of the field work on 16 May, the vessel flux-gate compass was calibrated in two exercises performed off Cannery Row. This enabled the vessel heading to be determined to within ±1 deg. The flux-gate compass was interfaced with the Trackpoint system (Sec. IV D 2), enabling the ROV position to be given in absolute rather than relative coordinates.

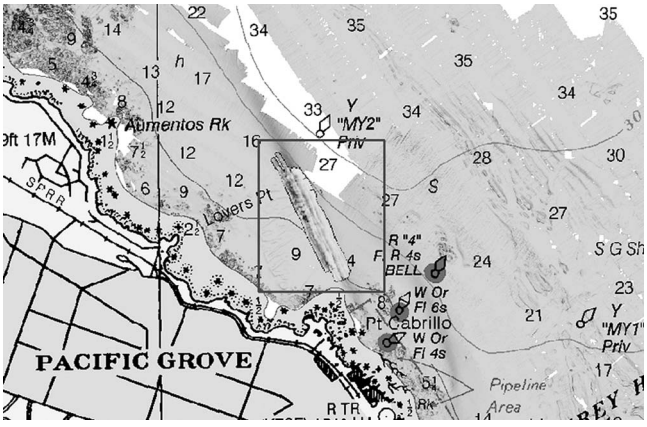


FIG. 2. Geographical region of the sidescan sonar survey off Pacific Grove in southern Monterey Bay on 15 May 2004, showing the high-frequency sidescan sonar mosaic.

The ROV was deployed with the vessel drifting or motoring gently to maintain a track parallel to the fault line. Three parallel track lines, located within the sidescan sonar survey area, were surveyed, with a nominal separation distance of 20 m. One of the track lines is shown in Fig. 3. Several large, dense aggregations of squid egg beds were circumnavigated in an attempt to map their extents, occasionally exceeding 10 m. The ROV was kept approximately 1–2 m from the bottom, with a forward/downward viewing angle of approximately 45 deg. Video from the ROV was recorded with the JVC unit on mini-digital videotape format. ROV positions were determined using a Trackpoint II+ ultra-short-baseline acoustic tracking system (ORE International, Houston, TX) and the vessel dGPS to yield real-world (x,y,z)-coordinates with ±2-m accuracy. The ROV depth was also determined by a pressure sensor mounted on the vehicle. ROV position, pressure-determined depth, and time in UTC were recorded onto the videotape using a Horita GPS-3 encoder.

Lone squids were observed in the vicinity of some egg beds. Occasional pairings and contested pairings were also observed, although the observations of spawning squid were incidental to the objective of surveying the egg beds.

VI. DATA POSTPROCESSING

Several data processing operations were undertaken in the Seafloor Mapping Laboratory immediately after completion of the cruises.

A. Postprocessing and mosaicking of sidescan sonar data

Digital processing and mosaicking of the 400-kHz sidescan sonar data were accomplished using the Isis Sonar and Delph Map software packages (Triton Elics International, Watsonville, California) and TNT Mips geographical information system (GIS) software (Microimages, Inc., Lincoln, Nebraska). Individual xtf track-line files were replayed, and bottom tracking of the sonar was supervised to aid in proper slant-range correction. Line files were snipped to remove portions with poor imagery from the beginning and/or end of the track line. Track lines were corrected for slant-range, layback, and lateral offset, and the position data for each line were smoothed using a speed filter. Each line was then gridded, geo-referenced, and exported from Isis Sonar/Delph Map in geo-referenced Tag(ged) Image File Format (GeoTIFF) (0.20-m and 0.10-m pixel size, UTM Zone 10, WGS84). Individual track-line TIFF images were imported into TNT Mips GIS software and areas of poor image quality were extracted and removed. Individual track lines were then overlaid to produce a mosaic image.

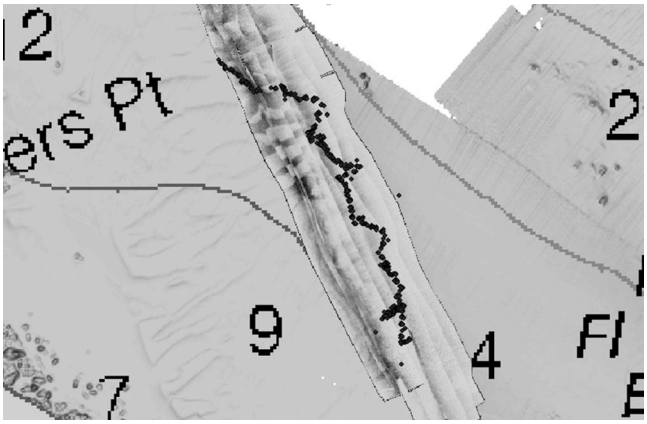


FIG. 3. First track of the ROV-borne video camera on 16 May 2004, superimposed on the sidescan sonar mosaic. The distance between the marked end points of the track is 603 m.

B. Video-data postprocessing

Video-data processing was completed using a JVC BR-DV600A mini-DV VCR with monitor display. Tapes were reviewed and positional data retrieved using the Horita GPS-3 decoder. The precise locations of squid egg bed observations were recorded in a text file, together with observational data regarding egg bed density, substrate type, and other characteristics. Information from the log sheets and text files was integrated into a database.

C. Entry of data into geographical information system (GIS)

Mosaic GeoTIFF images were exported from TNT Mips at 0.20- and 0.10-m resolutions into ArcGIS 8.3 software (ESRI, Redlands, California). An attribute table that included the parameters logged during the video-camera data analysis was also imported into the same GIS.

Use of ArcGIS enabled both the sonar data and video-camera tracks to be visualized. Individual or multiple sidescan sonar swaths could be displayed, as could video-camera tracks, at the command of the operator. In addition to simple visualization of data, GIS confers other benefits.[49] These include, for example, displaying the data at different scale sizes, expressing the cursor position in geographical coordinates, overlaying fields with a coordinate grid, measuring the distance between points, marking particular features with a polygon, among other things. Extracted images could also be exported for use outside the GIS.

VII. DATA ANALYSES

The first analysis involved presence–absence testing; the second, determination of the detection size limit; and the third, abundance estimation of squid eggs over an area.

A. Presence–absence testing

The primary goal of the testing was to prove, or verify, the association of characteristic features on the sidescan sonar images with the presence of squid egg beds, as well as the absence of squid egg beds in the absence of those same characteristic features. A secondary goal was to establish the size limit of detectability.

Test 1. By means of the ArcMap module within ArcGIS, the sidescan imagery was reviewed in fine detail by one of the co-authors (KF). Geographical areas with dapples or mottling believed characteristic of the presence of squid egg beds were identified. Areas lacking these features were also

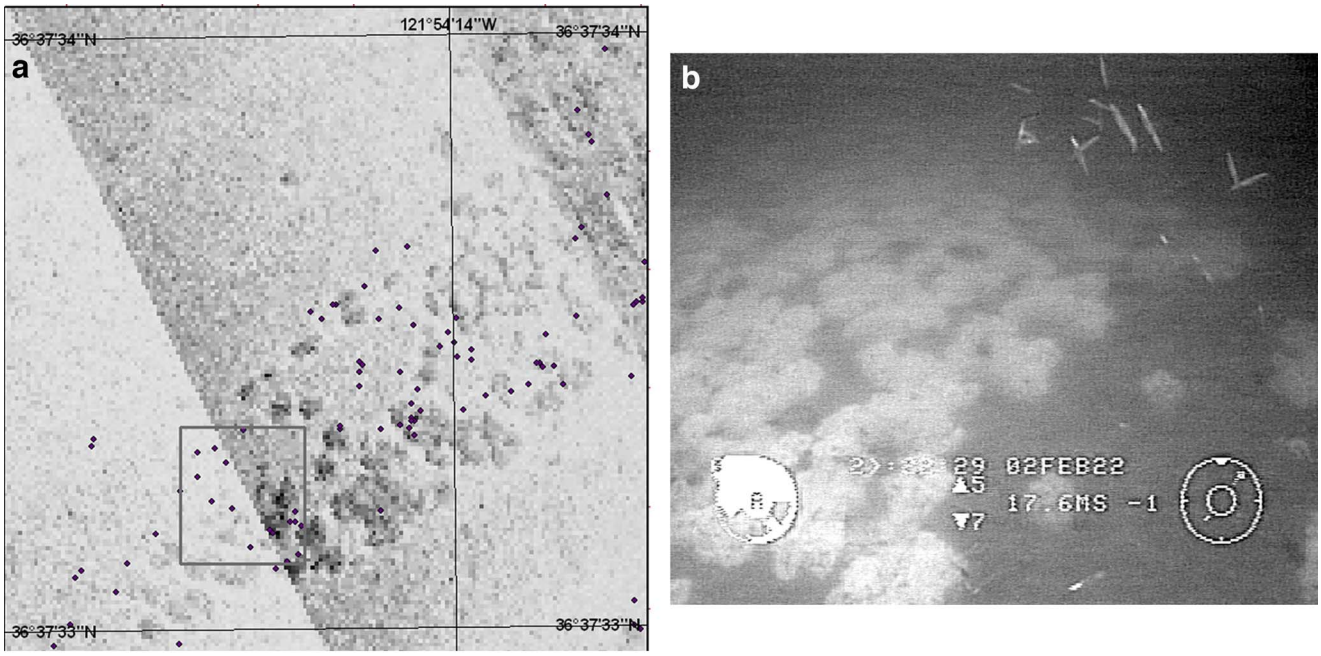


FIG. 4. Pair of images taken from the first test. Excerpt of sidescan sonar mosaic (a), with box indicating the region where the video-camera image (b) was captured. Adult squid are observed in the upper right corner of (b).

identified. A selected example of apparent egg presence is shown in Fig. 4(a). This area was also observed by the ROV-borne video camera. Time codes corresponding to the period that the ROV was inside the indicated box in Fig. 4(a) were determined. The particular time codes, together with time codes from another 29 selected regions of interest, were listed without reference to position or sidescan sonar observations to define a first test. These were used blindly by another co-author (RH) when examining the video-camera recordings to describe the presence or absence of squid egg beds. Related observations were made on the size, shape, and distribution of egg mops or beds or other benthos, e.g., kelp fronds, starfish, and sea stars, among other things.

Test 2. The process of formulating Test 1 was reversed in preparing a second blind test, which also endeavored to define the limits of detectability by sidescan sonar. An example of a video image selected for this test is shown in Fig. 5(a). The second co-author (RH) reviewed the ROV-borne video-camera tapes in detail, without reference to the sidescan sonar data, choosing a total of 30 representative regions of interest containing or lacking squid egg mops or beds. A list of the corresponding time codes was prepared without

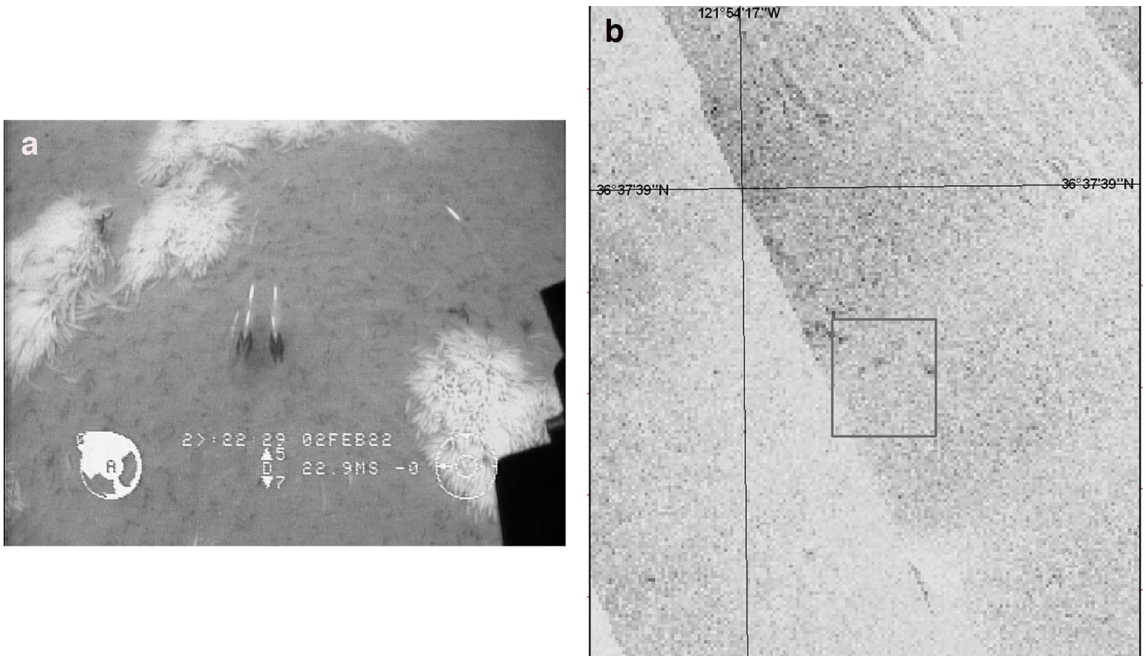


FIG. 5. Pair of images taken from the second test. Video–camera image (a) taken from the indicated area of the corresponding sidescan sonar mosaic excerpt (b). Two mating pairs and two loan squids are also visible in (a).

information on the video-based observations. The time codes were translated into geographical positions. The first coauthor (KF) then reviewed and interpreted the sidescan sonar imagery in a blind test to describe the markings as believed indicative of the presence or absence of squid egg beds.

The statistical significance of each test was gauged by means of both the binomial and Fisher exact test.[50]

B. Detection limit

The spatial resolution of the 400-kHz sidescan sonar array is about 10–20 cm over the range interval 12–23 m. Thus, there will be a lower limit to the size of egg beds that can be detected and identified. Determination of this limit was a secondary goal of the testing. Materials for the determination of the size limit were contained in the presence–absence testing. The associated video record was examined to establish the sizes of individual egg mops or beds when these were well separated from other clusters. Discrepancies in the statistical testing mentioned in the previous section were given particular attention.

C. Abundance estimation

A further analysis was performed to illustrate the quantification process. A quadrat of side length one arcsecond on each side, 30.9 m in the N–S direction and 24.8 m in the E–W direction, was defined, and the fraction of area covered by the squid egg beds was estimated. Assuming a mean diameter of egg capsules of 16 mm, based on several direct measurements, and mean number of eggs per capsule of 150, based on recent counts of eggs in sampled capsules, the total number, or abundance, of eggs in the quadrat was estimated.

VIII. RESULTS

Positions at which sidescan-sonar-based predictions of squid-egg presence or absence were subsequently, independently judged by reference to the video record are shown in Fig. 6. Agreement of the respective data is indicated by the use of open symbols; disagreement, by filled symbols.

Positions at which video observations of the presence or absence of squid eggs were subsequently, independently interpreted by the sidescan sonar images are shown in Fig. 7. As in the previous figure, agreement is indicated by open symbols and disagreement by filled symbols.

Both the binomial test and the Fisher exact test[50] were used to test the null hypothesis that the observed agreement in corresponding sonar and video images was due entirely to chance. Contingency tables are presented for the two tests in Tables I and II.

Each test involved 30 trials, each coincidentally yielding 24 positive outcomes in which the presence or absence of eggs was verified by the video record (Table I) or interpreted by the sonar imagery (Table II). The probability that 24 or more outcomes would have agreed was computed according to the binomial distribution assuming a probability of any single outcome being 0.5. The result was 0.7155×10^{-3} for each test.

The Fisher exact test was similarly applied to each set of data. The probability was computed of the distribution of outcomes shown in the respective tables together with their more extreme cases. The result for the first test (Table I) was 0.5939×10^{-4} and for the second (Table II), 0.4082×10^{-7}.

Thus, at least at the 0.001 level, no matter which test is used, the null hypothesis can be rejected. Characteristic features on the sidescan sonar images can be associated with the presence or absence of squid egg beds.

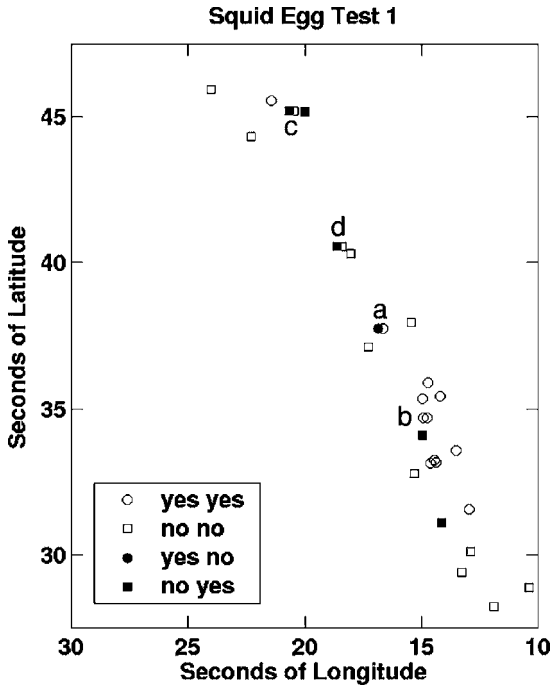


FIG. 6. Positions where sidescan-sonar-based predictions of squid-egg presence or absence were subsequently, independently judged by the video record. Agreement of the respective data is indicated by the use of open symbols; disagreement, by filled symbols.

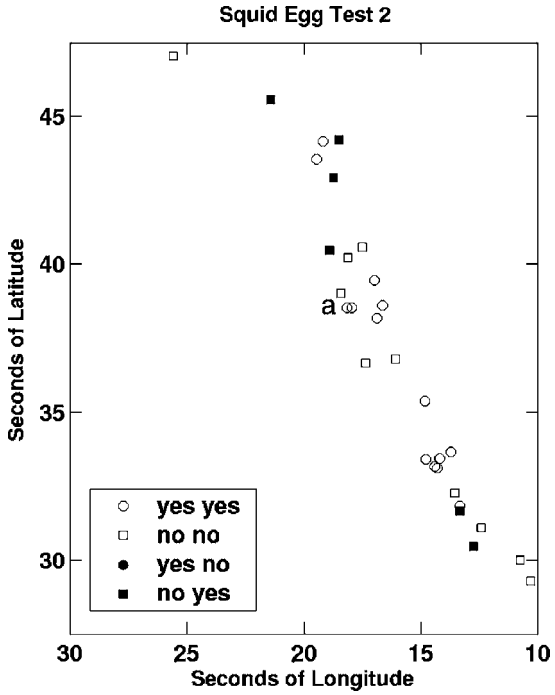


FIG. 7. Positions where video observations of squid-egg presence or absence were subsequently, independently interpreted by the sidescan sonar images. Agreement is indicated by open symbols; disagreement, by filled symbols.

TABLE I. Contingency table for Test 1, in which sidescan sonar predictions were made at selected positions, which were subsequently used to specify where to examine the video-camera record.

		Sidescan sonar		
		Presence	Absence	Sum
Video camera	Presence	12	5	17
	Absence	1	12	13
	Sum	13	17	30

A quadrat has been defined by the one-arc-second area at N 36° 37′33–34″ and W 121° 54′14–15″. The sidescan sonar image is shown in Fig. 8. The quadrat area is 764.6 m^2. The cumulative area covered by the squid egg beds according to the characteristic dapples was estimated by a graphical method as 49.0 m^2, or 6.4% of the whole. Assuming that the mean diameter of an egg capsule is 16 mm, which was measured for a limited number of retrieved capsules, the estimated number of capsules in the quadrat is 2.44×10^5. Based on an average number of eggs per capsule of 150, the estimated total number of eggs in the quadrat is 3.65×10^7.

TABLE II. Contingency table for Test 2, in which video-camera observations were made at selected positions, which were subsequently used to specify where to examine the sidescan sonar images.

		Video camera		
		Presence	Absence	Sum
Sidescan sonar	Presence	14	0	14
	Absence	6	10	16
	Sum	20	10	30

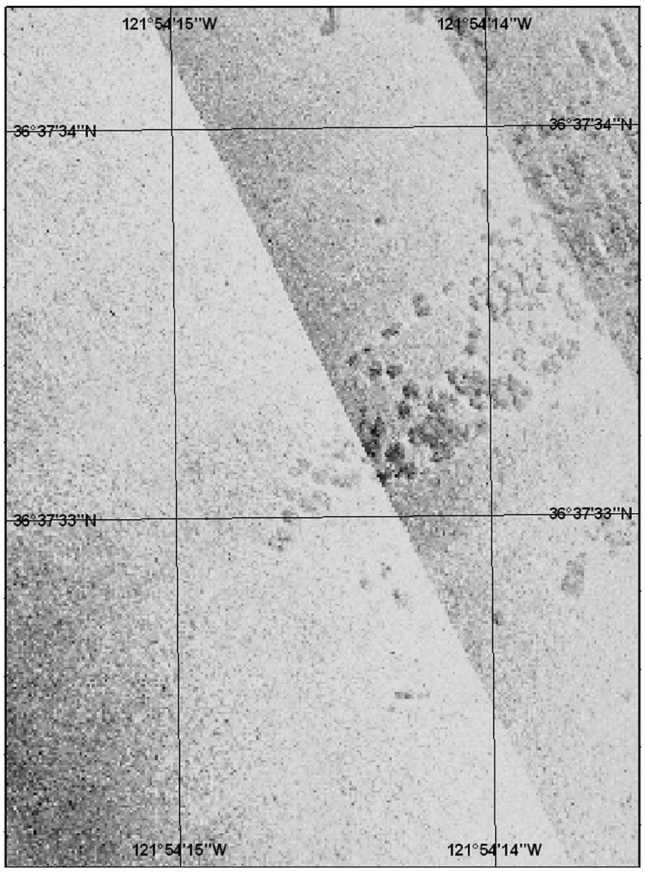


FIG. 8. Sidescan sonar image of a quadrat at N 36° 37′33–34″ and W 121° 54′14–15″ selected for estimating squid-egg abundance.

IX. DISCUSSION

A. Statistical tests

Characteristic dapple or mottling on the sidescan sonar images at 100 and 400 kHz was first associated with squid egg beds during a cruise in Monterey Bay in May 2003. Data collected one year later, which are analyzed in this work only at the higher frequency, verify the association. The visual images in Figs. 4 and 5 are clearly suggestive, but the robustness of the association has been statistically established through two blind tests.

In the first test, predictions of the presence or absence of squid egg beds were made solely on the basis of the sidescan sonar data, then tested by reference to the video camera recordings. In the second test, areas observed to contain or lack egg beds were used to specify regions of interest in the sidescan sonar images requiring interpretation. Each test was prepared by a co-author and taken by another co-author without knowing the initial sonar interpretation or video observation. The null hypothesis, that the apparent agreement of predictions or interpretations with observations was due entirely to chance, was tested by each of two tests. The null hypothesis was rejected at the 0.001 confidence level by the binomial test and the 0.0001 confidence level by the Fisher exact test.

It is illuminating to examine the failed examples in the two tests. In the first test, in which sonar images were used to specify positions for examination of the video record, there were a total of six failures. Comments made during review of the video record in Test 1 are listed for each instance of disagreement. In only one case did the sidescan sonar image indicate the presence of eggs that was not verified by the video record. In the region of pair a, eggs were not observed along the third ROV track, although their presence was confirmed by the second ROV track. This probably represents a simple miss due to the video camera being oriented in the wrong direction when passing the egg beds. In the other five cases of disagreement, the sidescan sonar images did not indicate the presence of eggs, which were subsequently observed on the video record. The following comments are arranged in order of position from north to south. (i) Two very small mops were observed along the second ROV track, but not along the first ROV track, indicated in pair c. (ii) One tiny egg mop was seen along the first ROV track at the edge of the region of interest. (iii) A few very small, widely dispersed mops were seen along the third ROV track, but not along the second ROV track, in pair d. (iv) A few large clumps, *ca.* 1–3 m, were observed along the third ROV track at the edge of the region of interest. (v) Two small patches were observed along the third ROV track. In four of the five cases here, the eggs occurred in small mops of diameter less than 0.5 m.

In the second test, there were no instances when the video record indicated the absence of eggs and the sidescan

sonar imagery suggested their presence. There were six instances when the sidescan sonar imagery failed to register the presence of eggs that had been seen initially on the video record. The following comments were made on these instances during preparation of Test 2: 4–6 small patches *ca.* 0.5 m each; 1–2 very small egg patches; linear string of eggs; 1 patch *ca.* 10 capsules; 2 patches *ca.* 1 m; and 1 small patch. In four of these cases, the egg mops were, at most, 0.5 m in diameter. In the other cases, the linear string of eggs was missed, possibly because of its unusual shape; and in the other case, of two patches *ca.* 1 m each, possibly because of inaccurate positioning.

B. Size detectability

In view of the failed examples in the two tests, it appears that egg mops smaller than about 0.5 m cannot be detected with the system as currently configured and used. This approximate detection limit applies to the use of a 400-kHz sidescan sonar of 0.5-deg beamwidth when towed at 3-knot speed with a ping rate of 13/s. The resolution is about 10–20 cm out to about the 25-m range on either side of the towfish. Thus, an object of 50-cm diameter will be spanned by about 3–5 pixels, with corresponding size in the image. This number is evidently too small to resolve squid egg beds.

The present analysis has only attempted to distinguish the presence and absence of squid eggs. Observed aggregations of eggs ranged in size from roughly circular mops smaller than 0.5 m to much larger, irregularly shaped clusters of beds. Two such aggregations were measured according to their sidescan sonar image; one was roughly elliptical in shape, with axes 2.7 and 7.4 m; another was cross-shaped, with similar 5-m axes. Egg mops greater than 0.5 m and egg beds greater than 1 m have been detected by the sidescan sonar imagery.

Other organisms were also observed in the video record. Examination of the sidescan sonar images at the corresponding locations has failed to reveal identified individual drift kelp fronds (*Macrocystis pyrifera*) and sea stars, e.g., *Pycnopodia helianthoides* and *Asterina miniata*. It is appreciated that these were typically below the detectability size limit, although some kelp strands and the largest *Pycnopodia* sea stars exceeded 0.5 m and still did not appear as distinctive features in the acoustic records. The low vertical relief of these targets was probably the primary reason that the sidescan sonar failed to effectively image them.

C. Positioning issues when comparing diverse data

Comparison of the sonar and video data, which were collected on different days from different submerged platforms, has also required a closer look at the issues of positioning accuracy and precision. While differential GPS was used, this is updated only at intervals, with an intrinsic accuracy of about ±1–2 m in the Monterey Bay region. The accuracy in positioning the sidescan sonar towfish (Sec. IV A 1) is poorer since it is towed, and its position must be estimated from the amount of cable deployed and its depth. Further, the towfish is subject to accelerations due to wave action on the towing vessel, which can change the velocity and position of the towfish relative to the dGPS receiver. Making allowance for this, inaccuracies in towfish positioning can be ±5 m, depending on the exact circumstances of towing, including sea state, currents, and direction of ship movement. The accuracy in positioning the ROV-borne video camera is better than that of the towfish, for the ROV is decoupled from vessel motions by having a slack tether, and the ROV is acoustically tracked from the vessel. Thus the positions of the towfish and ROV-borne video camera can each be known to within about ±1–2 m under good conditions, but closer to ±5 m under poorer conditions, as with a fresh breeze. In addition, the camera height and orientation influences the field of view, which can differ from the position itself by several meters. Under the worst circumstances, the combined uncertainty of these two can be as much as ±10 m. In comparing the sonar and video observations, agreement was sought to within 3–5 m, occasionally up to about 6–7 m.

Performance of mosaicking also requires comparison of data from transect to transect, if of similar origin. Clues to positioning inaccuracy are evident in the sidescan sonar record. Distortions may be noticed, as in stretched egg bed features in some images. The process of mosaicking can be improved, but the experience of fine-scale optical mosaicking in the laboratory indicates that this is an intricate problem that is inherently three-dimensional in character.[51] The use of terrain constraints can significantly improve underwater mosaicking, whether optical or acoustical, if the number of images to be combined is not excessive (C. Roman, pers. comm.).

D. An abundance estimation

The abundance of squid eggs was estimated for a particular quadrat, yielding the number 3.65×10^7. This quantity is not dissimilar from an estimation performed in a different quadrat the preceding year. This is the first numerical estimate of egg abundance over a region of interest by a rapid, remote, synoptic, noninvasive method. Extension of the quadrat-based estimation to a larger survey area is straightforward.

The authors know of no other acoustic estimate of the abundance of benthic eggs of any squid species. The authors know of two studies reporting abundance estimates of benthic eggs of squid species, but based on SCUBA-diver surveys using principles of random or nonrandom sampling. In a study by Sauer *et al.*,[44] two basic abundance estimates of eggs of *Loligo vulgaris reynaudii* were derived, distinguished by the respective statistical analyses: 3.162×10^9 and 1.290×10^8 eggs over respective areas of 74.36 and 1.646 km^2. Corresponding numerical densities were 42.52 eggs/m^2 in the larger area and 78.32 eggs/m^2 in the smaller area. In a study by Moltschaniwskyj and Pecl,[52] abundance estimates of benthic eggs of southern calamary (*Sepioteuthis australis*) were derived for two different bays in each of two years. The numbers for one bay were 7.68×10^6 and 1.14×10^7 eggs over respective areas of 3.67 and 4.19 ha, and 2.47×10^6 and 1.43×10^6 eggs over respective

areas of 1.35 and 1.36 ha. The corresponding numerical densities were 209 and 272 eggs/m^2 for the first bay and 182 and 105 eggs/m^2 for the second. These numbers are compared with the numerical density of *Loligo opalescens* eggs determined by the acoustic method reported here, namely 47 720 eggs/m^2, that is, about three orders of magnitude greater, although estimated over an area that is less than 1/1000 of the areas in the first study and 1/10 of the areas in the second study. Another major difference is represented by the relative areal coverage. This was partial in the SCUBA-diver surveys, but total in the present acoustic survey using sidescan sonar.

E. Future work

It should be possible to resolve squid egg beds of dimensions smaller than 0.5 m, as with a more directional, higher-frequency sidescan sonar, with a higher ping rate and a smaller maximum range than was used in the reported work. This would be very useful for extending the detection and identification capability of the method, hence its quantification capability as well. Distinguishing other benthic organisms, e.g., kelp and starfish, would also be useful.

It would be interesting and potentially useful to perform sidescan sonar surveying from other platforms, especially from autonomous underwater vehicles (AUVs). The example of the Remotely Environmental UnitS (REMUS) AUV[53] comes to mind as a particularly attractive vehicle because it carries sidescan sonars on port and starboard sides as standard equipment. The operating frequency of these is either 600 or 900 kHz.

Automatic recognition of squid egg beds on the sidescan sonar image should be possible, as should computation of areas covered by the egg beds. However, can this be done reliably, accurately and with minimal operator intervention?

Combining observation of squid egg beds with that of adult squid in the water column is also important. Acoustic measurement of squid in the water column is eminently feasible, as demonstrated through a number of works cited in Sec. I B. In this way, local population recruitment could be measured directly, and potential future recruitment could be estimated by quantification of the egg beds. The complex of interactions that govern the success of recruitment, including the known attractive effects of egg beds, could help gauge these components as well as others used by fishery managers and population biologists.

It is expected that some basic questions on the biology of *Loligo opalescens* may be answered through the extended use of acoustic methods. These include, for example, questions on the geographical distribution of egg beds and their association with the benthic habitat. Statistical properties of aggregation can be determined by application of the variogram, the basic structural tool of geostatistics.[54,55] The potential of acoustics for monitoring extent and rate of egg deposition on the squid spawning grounds is apparent, especially with regard to seasonal and yearly fluctuations that could be correlated eventually with fishery catch records of adults. An example of this approach is reported by Moltschaniwskyj and Pecl,[52] mentioned in Sec. IX D, who used SCUBA in shallow water to count small egg cases of *Sepioteuthis* in sea grass. The approach may serve the needs of both managers responsible for the squid fishery and scientists trying to understand the dynamics of squid spawning during an ongoing fishery.

It is a useful exercise to estimate the degree of acoustic coverage that can be achieved by sidescan sonar in a period of hours. If the full track width of sidescan sonar coverage is 100 m and 100% overlap is assumed on adjacent, port or starboard parallel tracks, then at a vessel speed of 3 knots, a 5.4-km long transect of effective area 0.27 km^2 can be covered in 1 h, 2.12 km^2 in 8 h. This is a sizable fraction of the total estimated area of the primary spawning grounds in southern Monterey Bay, namely 10 km^2.

A fundamental acoustic problem is also waiting to be solved, namely that of determining the backscattering cross section of individual squid egg capsules and how this changes with development phase. This problem might be approached through numerical solution of the wave equation by the finite-element method[56,57] but as applied to acoustic scattering problems, as in the boundary-element method, e.g., Ref. 58, or infinite-element method, e.g., Ref. 59. Data on the physical properties of individual egg capsules, namely mass density and compressibility, would be required. The scattering problem could also be addressed by measurement, but under controlled conditions.

A second outstanding acoustic problem is modeling scattering from an egg mop or bed. The morphology of the upper surface of a mop or bed is known to be knobbly, with an amplitude of unevenness of order of 1 cm. Interstices are present between neighboring egg capsules. Egg mops and beds may also bend under the influence of currents, including surface-wave action. An appropriate theory of acoustic scattering by such a surface may resemble that for sound scattering by sea ice dendrites at the air–water interface.[60]

Development of a method for calibrating sidescan sonar would also be useful. If the standard-target method[61–63] could be extended to sidescan sonar, as has recently been accomplished for multibeam sonar,[64] then the measurements could be performed in an absolute sense. Potential benefits would include demonstrable data quality and maintenance of the same over time, as well as a capability for the intercomparison of data collected with different sidescan sonars.

X. CONCLUSIONS

(1) Egg beds of *Loligo opalescens* in Monterey Bay can be detected, identified, and quantified by high-frequency sidescan sonar imagery.

(2) The lower size limit of detectability of squid egg beds by a 400-kHz sidescan sonar is about 0.5 m.

(3) The method is ready for monitoring applications.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of many colleagues and organizations. C. Gubala (JC Headwaters Canada) is thanked for lending and operating an EY echo sounder during the first, laboratory-based investigation. A. Henry (California Department of Fish and Game) is thanked

for valuable support, participation in the first cruise, and facilitating use of R/V Mako, with the following officers and crew: Captain M. Kibby, Engineer R. L. Michalski, R. Hornady, J. DeWitt, T. Tanaka. The following are also thanked for their participation in this first cruise: C. Chan and T. Wadsworth (CDFG); and D. Sullivan, A. Hochstaedter, Y. Yokozawa, and J. Zande (Monterey Peninsula College). Contributions by the following students are acknowledged: A. Rupp (California State University Monterey Bay), and G. Dayhuff, M. Robinson, J. Kibele, D. Williams, S. Okano, and N. Capps (MPC). The following SCUBA divers are thanked for their vital underwater labors: R. Hornady, H. King, C. Dawson, and D. Osorio (CDFG); and S. Clabuesch (University of California at Santa Cruz). G. Shipton (Triton Elics International) is thanked for preparing the sidescan sonar, and J. Condiotty (Simrad) for arranging a loan of the EK60 echo sounder. M. Windham (CDFG) is thanked for logistical support. S. Fangman, NOAA Channel Islands National Marine Sanctuary, is thanked for valuable support, including facilitating the use of R/V Shearwater during the second cruise, with NOAA captains L. Moody and T. Shinn, as well as personal participation during the cruise. CDFG is again thanked for allowing the participation of R. L. Michalski, T. Bishop, K. M. O'Brien, L. M. Laughlin, and D. Porzio during the second cruise. D. Vasey, Santa Barbara City College, is thanked for vital scientific ROV support, as well as allowing the following students to impart their general seagoing expertise in the scientific operations of the cruise: D. Adams, A. Sensabaugh, J. Valenzuela, and R. Remley. M. Parmenter, C. Sellers, and K. Buresch are all thanked for contributions to the figures. The authors are grateful for funding from the National Sea Grant, Essential Fish Habitat Program, and especially the interest of program manager E. Anderson, Sea Grant Project No. NA16RG2273. This is Woods Hole Oceanographic Institution contribution number 11372.

[1]P. R. Boyle and S. Boletzky, "Cephalopod populations: definitions and dynamics," Philos. Trans. R. Soc. London, Ser. B **351**, 985–1002 (1996).
[2]G. V. Morejohn, J. T. Harvey, and L. T. Krasnow, "The importance of *Loligo opalescens* in the food web of marine vertebrates in Monterey Bay, California," in *Biological, Oceanographic, and Acoustic Aspects of the Market Squid, Loligo opalescens Berry*, edited by C. W. Recksiek and H. W. Frey, Calif. Dep. Fish Game Fish Bull. **169**, 67–98 (1978).
[3]W. G. Fields, "The structure, development, food relations, reproduction, and life history of the squid *Loligo opalescens* Berry," Fish. Bull. **131**, 1–108 (1965).
[4]C. M. Dewees and R. J. Price, "Overview of the squid fishery on the Pacific coast of the United States," in *Proceedings of the International Squid Symposium*, 9–12 August 1981, *Boston, Massachusetts* (New England Fisheries Development Found., Boston, 1982), pp. 197–212.
[5]M. Vojkovich, "The California fishery for market squid (*Loligo opalescens*)," CalCOFI Rep. **39**, 55–60 (1998).
[6]California Department of Fish and Game, "Market squid fishery management plan: draft," State of California, The Resources Agency, Department of Fish and Game, Marine Region, 7 July 2003.
[7]B. J. Macewicz, J. R. Hunter, N. C. H. Lo, and E. L. LaCasella, "Fecundity, egg deposition, and mortality of market squid (*Loligo opalescens*)," Fish. Bull. **102**, 306–327 (2004).
[8]W. H. H. Sauer, M. J. Smale, and M. R. Lipinski, "The location of spawning grounds, spawning and schooling behaviour of the squid *Loligo vulgaris reynaudii* (Cephalopoda: Myopsida) off the Eastern Cape Coast, South Africa," Mar. Biol. (Berlin) **114**, 97–107 (1992).
[9]J. K. T. Brodziak and A. A. Rosenberg, "A method to assess squid fisheries in the north-west Atlantic," ICES J. Mar. Sci. **50**, 187–194 (1993).
[10]G. J. Pierce, N. Bailey, Y. Stratoudakis, and A. Newton, "Distribution and abundance of the fished population of *Loligo forbesi* in Scottish waters: analysis of research cruise data," ICES J. Mar. Sci. **55**, 14–33 (1998).
[11]D. J. Agnew, R. Baranowski, J. R. Beddington, S. des Clers, and C. P. Nolan, "Approaches to assessing stocks of *Loligo gahi* around the Falkland Islands," Fish. Res. **35**, 155–169 (1998).
[12]S. J. Walsh, P. A. Koeller, and W. D. McKone, eds. *Proceedings of the International Workshop on Survey Trawl Mensuration. Northwest Atlantic Fisheries Centre, St. John's, Newfoundland, March 18–19, 1991*, Canadian Tech. Rep. Fish. Aquat. Sci. **1911**, 1–114 (1993).
[13]D. R. Gunderson, *Surveys of Fisheries Resources* (Wiley, New York, 1993).
[14]T. Okutani and J. McGowan, "Systematics, distribution, and abundance of the epiplanktonic squid (Cephalopoda, Decapoda) larvae of the California Current, April, 1954–March, 1957," Bull. Scripps. Inst. Oceanogr. **14**, 1–90 (1969).
[15]M. R. Maxwell, A. Henry, C. D. Elvidge, J. Safran, V. R. Hobson, I. Nelson, B. T. Tuttle, J. B. Dietz, and J. R. Hunter, "Fishery dynamics of the California market squid (*Loligo opalescens*), as measured by satellite remote sensing," Fish. Bull. **102**, 661–670 (2004).
[16]M. Clarke, L. Allcock, and M. B. Santos, "Estimating cephalopod biomass: workshop report," Bull. Mar. Sci. **71**, 47–65 (2001).
[17]G. M. Cailliet and D. L. Vaughan, "A review of the methods and problems of quantitative assessment of *Loligo opalescens*," Biol. Oceanogr. **2**, 379–400 (1983).
[18]K. G. Foote, "Quantitative fisheries research surveys, with special reference to computers," in *Computers in Fisheries Research*, edited by B. A. Megrey and E. Moksness (Chapman and Hall, London, 1996), pp. 80–112.
[19]R. M. Starr and R. E. Thorne, "Acoustic assessment of squid stocks," in *Squid Recruitment Dynamics. The Genus Illex As a Model, the Commercial Illex Species and Influences on Variability*, edited by P. G. Rodhouse, E. G. Dawe, R. K. O'Dor, FAO Fish. Tech. Pap. **376**, 181–198 (1998).
[20]C. Goss, D. Middleton, and P. Rodhouse, "Investigations of squid stocks using acoustic survey methods," Fish. Res. **54**, 111–121 (2001).
[21]M. J. Roberts, M. Barange, M. R. Lipinski, and M. R. Prowse, "Direct hydroacoustic observations of chokka squid *Loligo vulgaris reynaudii* spawning activity in deep water," S. Afr. J. Mar. Sci. **24**, 387–393 (2002).
[22]A. Kawabata, "Target strength measurements of suspended live ommastrephid squid, *Todarodes pacificus*, and its application in density estimations," Fish. Sci. **71**, 63–72 (2005).
[23]D. L. Vaughan and C. W. Recksiek, "An acoustic investigation of market squid, *Loligo opalescens*," in *Biological, Oceanographic, and Acoustic Aspects of the Market Squid, Loligo opalescens Berry*, edited by C. W. Recksiek and H. W. Frey, Calif. Dep. Fish Game Fish Bull. **169**, 135–147 (1978).
[24]D. L. Vaughan and C. W. Recksiek, "Detection of market squid, *Loligo opalescens*, with echo sounders," CalCOFI Rep. **20**, 40–50 (1979).
[25]K. Jefferts, J. Burczynski, and W. G. Pearcy, "Acoustical assessment of squid (*Loligo opalescens*) off the Central Oregon coast," Can. J. Fish. Aquat. Sci. **44**, 1261–1267 (1987).
[26]J. E. Ehrenberg, "A comparative analysis of in situ methods for directly measuring the acoustic target strength of individual fish," IEEE J. Ocean. Eng. **OE-4**(4), 141–152 (1979).
[27]C. W. Recksiek and H. W. Frey, in *Biological, Oceanographic, and Acoustic Aspects of the Market Squid, Loligo opalescens Berry*, Calif. Dep. Fish Game Fish Bull. **169** (1978).
[28]R. F. Hixon, "*Loligo opalescens*," in *Cephalopod Life Cycles, Vol. 1. Species Accounts*, edited by P. R. Boyle (Academic, London, 1983), pp. 95–114.
[29]G. D. Jackson, "Statolith age estimates of the loliginid squid *Loligo opalescens* (Mollusca: Cephalopoda): Corroboration with culture data," Bull. Mar. Sci. **54**, 554–557 (1994).
[30]J. Butler, D. Fuller, and M. Yaremko, "Age and growth of market squid (*Loligo opalescens*) off California during 1998," CalCOFI Rep. **40**, 191–195 (1999).
[31]L. D. Zeidberg and W. M. Hamner, "Distribution of squid paralarvae, *Loligo opalescens* (Cephalopoda: Myopsida), in the southern California Bight in the three years following the 1997–1998 El Nino," Mar. Biol. (Berlin) **141**, 111–122 (2002).
[32]E. S. Hobson, "Spawning in the Pacific coast squid *Loligo opalescens*," Underwat. Nat. **3**, 20–21 (1965).
[33]A. C. Hurley, "Mating behavior of the squid *Loligo opalescens*," Mar. Behav. Physiol. **4**, 195–203 (1977).
[34]R. T. Hanlon, "Mating systems and sexual selection in the squid *Loligo*:

How might commercial fishing on spawning squids affect them?" CalCOFI Rep. **39**, 92–100 (1998).
[35]R. R. Leos, "The biological characteristics of the Monterey Bay squid catch and the effect of a two-day-per-week fishing closure," CalCOFI Rep. **39**, 204–211 (1998).
[36]J. W. Forsythe, N. Kangas, and R. T. Hanlon, "Does the California Market Squid, *Loligo opalescens*, spawn naturally during the day or at night? A note on the successful use of ROVs to obtain basic fisheries biology data," Fish. Bull. **102**, 389–392 (2004).
[37]R. T. Hanlon, N. Kangas, and J. W. Forsythe, "Egg capsule deposition and how behavioral interactions influence spawning rate in the squid *Loligo opalescens* in Monterey Bay, California," Mar. Biol. (Berlin) **145**, 923–930 (2004).
[38]A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams* (Wiley, Chichester, 2000).
[39]J. A. McGowan, "Observations on the sexual behavior and spawning of the squid, *Loligo opalescens*, at La Jolla, California," Calif. Fish and Game **40**, 47–54 (1954).
[40]L. D. Zeidberg, W. Hamner, K. Moorehead, and E. Kristof, "Egg masses of *Loligo opalescens* (Cephalopoda: Myopsida) in Monterey Bay, California following the El Nino event of 1997–1998," Bull. Mar. Sci. **74**, 129–141 (2004).
[41]P. G. Rodhouse, E. G. Dawe, and R. K. O'Dor, in *Squid Recruitment Dynamics. The Genus Illex as a Model, the Commercial Illex Species and Influences on Variability*, FAO Fish. Tech. Pap. **376** (1998).
[42]J. P. Fish and H. A. Carr, *Sound Underwater Images, A Guide to the Generation and Interpretation of Side Scan Sonar Data*, 2nd ed. (Lower Cape Publishing, Orleans, MA, 1990).
[43]J. P. Fish and H. A. Carr, *Sound Reflections, Advanced Applications of Side Scan Sonar* (Lower Cape Publishing, Orleans, MA, 2001).
[44]W. H. H. Sauer, C. McCarthy, M. J. Smale, and A. S. Koorts, "An investigation of the egg distribution of the chokka squid, *Loligo vulgaris reynaudii*, in Krom Bay, South Africa," Bull. Mar. Sci. **53**, 1066–1077 (1993).
[45]L. N. Andersen, "The new Simrad EK60 scientific echo sounder system," J. Acoust. Soc. Am. **109**, 2336 (2001).
[46]H. Bodholt, H. Nes, and H. Solli, "A new echo-sounder system," Proc. Inst. Acoust. **11**(3), 123–130 (1989).
[47]K. G. Foote and G. Stefánsson, "Definition of the problem of estimating fish abundance over an area from acoustic line-transect measurements of density," ICES J. Mar. Sci. **50**, 369–381 (1993).
[48]H. R. Skjoldal, P. H. Wiebe, and K. G. Foote, "Sampling and experimental design," in *ICES Zooplankton Methodology Manual*, edited by R. Harris, P. Wiebe, J. Lenz, H. R. Skjoldal, and M. Huntley (Academic, San Diego, 2000), pp. 33–53.
[49]J. Breman, ed. *Marine Geography: GIS for the Oceans and Seas* (ESRI Press, Redlands, CA, 2002).
[50]J. H. Zar, *Biostatistical Analysis*, 4th ed. (Prentice-Hall, Upper Saddle River, NJ, 1999).
[51]R. Preston, "Capturing the unicorn," The New Yorker **81**(8), 28–33 (2005).
[52]N. A. Moltschaniwskyj and G. T. Pecl, "Small-scale spatial and temporal patterns of egg production by the temperate lolilginid squid *Sepioteuthis australis*," Mar. Biol. (Berlin) **142**, 509–516 (2003).
[53]B. R. Allen, R. Stokey, T. Austin, N. Forrester, R. Goldsborough, M. Purcell, and C. von Alt, "REMUS: a small, low cost AUV; system description, field trials and performance results," *Proc. MTS/IEEE Oceans'97 Conf.*, Vol. **2**, pp. 994–1000 (1997).
[54]N. A. C. Cressie, *Statistics for Spatial Data* (Wiley, New York, 1991).
[55]J. Rivoirard, J. Simmonds, K. G. Foote, P. Fernandes, and N. Bez, *Geostatistics for Estimating Fish Stock Abundance* (Blackwell Science, Oxford, 2000).
[56]O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 5th ed. (Butterworth Heinemann, Oxford, 2000).
[57]D. S. Burnett, *Finite Element Analysis: from Concepts to Applications* (Addison-Wesley, Reading, MA, 1988).
[58]D. T. I. Francis, "A gradient formulation of the Helmholtz integral equation for acoustic radiation and scattering," J. Acoust. Soc. Am. **93**, 1700–1709 (1993).
[59]D. S. Burnett, "A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion," J. Acoust. Soc. Am. **96**, 2798–2816 (1994).
[60]T. K. Stanton, K. C. Jezek, and A. J. Gow, "Acoustical reflection and scattering from the underside of laboratory grown sea ice: Measurements and predictions," J. Acoust. Soc. Am. **80**, 1486–1494 (1986).
[61]K. G. Foote, "Optimizing copper spheres for precision calibration of hydroacoustic equipment," J. Acoust. Soc. Am. **71**, 742–747 (1982).
[62]K. G. Foote, "Maintaining precision calibrations with optimal copper spheres," J. Acoust. Soc. Am. **73**, 1054–1063 (1983).
[63]K. G. Foote, H. P. Knudsen, G. Vestnes, D. N. MacLennan, and E. J. Simmonds, "Calibration of acoustic instruments for fish density estimation: a practical guide," ICES Coop. Res. Rep. **144** (1987), 69 pp.
[64]K. G. Foote, D. Chu, T. R. Hammar, K. C. Baldwin, L. A. Mayer, L. C. Hufnagle, Jr., and J. M. Jech, "Protocols for calibrating multibeam sonar," J. Acoust. Soc. Am. **117**, 2013–2027 (2005).

Separation of thermoviscous losses in Celcor™ ceramic[a)]

Jin Liu and Steven L. Garrett[b)]
Graduate Program in Acoustics, The Pennsylvania State University Applied Research Laboratory, State College, Pennsylvania 16804

Gregory S. Long[c)] and Ayusman Sen[d)]
Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 29 July 2005; revised 21 November 2005; accepted 25 November 2005)

A series of three papers by Sabatier and coauthors reported successive improvements to both the theory and measurement of excess attenuation due to the wall porosity of Celcor™. We report an independent measurement of excess attenuation due to wall porosity of Celcor conducted to study the attenuation due to viscous and thermal effects separately for a thermoacoustic refrigeration application. A 38-mm-long sample was placed at the center of a 700-mm-long electrodynamically driven plane wave resonator. The location of the Celcor within the standing wave would select different contributions from viscosity and thermal conductivity. As per the technique described by Moldover *et al.* [Modern Acoustical Techniques for the Measurement of Mechanical Properties (Academic, San Diego, 2001)], the quality factors Q were measured between 230 Hz and 1.6 kHz in air at atmospheric pressure on a pristine sample. Wall pores were then blocked by an *in situ* polymerization sealing process. There was no observed decrease in attenuation of even modes (dominated by thermal loss) for the coated sample below 1.0 kHz. The volume exclusion due to pore sealing was inferred by the increase in even mode resonance frequencies in good agreement with other determinations of wall porosity. Thermoviscous losses were computed with DELTAE and compared to measured losses. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159291]

PACS number(s): 43.35.Ud, 43.20.Mv, 43.20.Ks [MFH] Pages: 857–862

I. INTRODUCTION

The use of Celcor™ as a thermoacoustic stack material was first suggested to us by scientists at the National Center for Physical Acoustics (NCPA). Due to a long-standing interest at NCPA in the acoustics of porous media, particularly soils, several measurements of the effects of the porosity of the ceramic cell walls were undertaken. In thermoacoustic applications, Celcor has been an attractive stack material due to its regular pore geometry, ease of shaping, low cost, and availability. The Celcor ceramic material has been used as the stack material for several standing-wave thermoacoustic devices.[1–3] Due to the wall porosity, as sketched in Fig. 1, the possibility of excess attenuation due to the wall porosity was a cause for concern.

Theory and experiments were reported for sound propagation in porous media consisting of straight capillary tubes having square cross sections. The theory[4] assumed a nonporous wall. A single microphone measurement method was used to determine the propagation constants in the square-pore media. A tortuosity factor of 1.1 was chosen to improve the agreement between theory and experiment. The measured absorption coefficient exhibited good agreement with the theory at low frequencies.

Arnott, Sabatier, and Raspet[5] enhanced this theory by considering the wall pores. The continuity equation was modified to include the transverse velocity gradient due to the wall pores. Density changes of the gas in the wall pores are assumed to be isothermal. Using this model, the agreement between theory and measurement was improved both in attenuation and phase velocity and no tortuosity correction was needed. But at high frequency, the predicted attenuation was still observed to be lower than the measurement. Bernard, Velea, and Sabatier[6] reported a method to seal the wall pores by coating them with a very thin layer of polyurethane varnish diluted with equal parts of mineral spirits and verified the application of the theory of sound propagation in square tube pores sample at frequencies above 1500 Hz.

The absorption coefficient is a combined effect of both thermal and viscous losses. In a standing-wave thermoacoustic refrigerator,[7] the position of the stack in the standing wave field determines how much of the losses will be due to viscosity and how much is due to thermal relaxation effects between the gas and the stack. In typical electrically driven standing wave thermoacoustic refrigerators,[8,9] the stack is located much closer to a pressure antinode than to a velocity antinode, so the effects of thermal relaxation dominate the losses due to shear viscosity.

Because thermal relaxation loss dominates viscous loss in a practical standing-wave thermoacoustic machine, interest in "excess" attenuation due to the Celcor wall porosity is not merely academic. The apparatus and measurements described in this paper were undertaken to determine whether

[a)]Portions of this work were presented at the 143rd Meeting of the Acoustical Society of America held in Pittsburgh, PA on 6 June 2002 and the abstract for that talk was printed in the Journal of the Acoustical Society of America, volume 111, No. 5, Part 2, May, 2002.
[b)]Electronic mail: sxg185@psu.edu
[c)]Currently at Los Alamos National Lab. Electronic mail: glong@lanl.gov
[d)]Electronic mail: asen@psu.edu

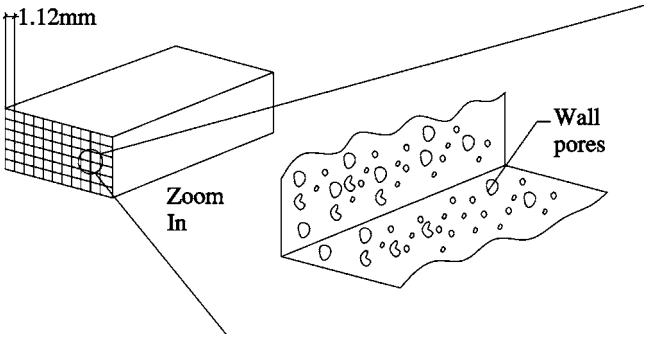


FIG. 1. Schematic representation of the cell structure of Celcor™ ceramic with an expanded view of a portion of cell wall illustrating the porous nature of the ceramic wall surface. Photomicrographs of actual cell walls are shown in Fig. 2.

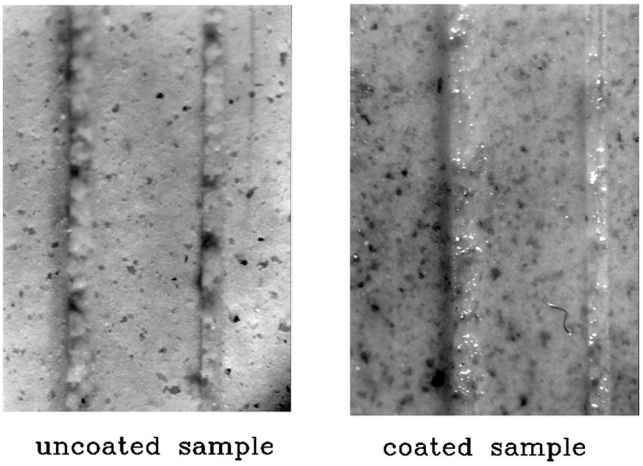


FIG. 2. Comparison of the surface of the walls of the pores before and after coating using the *in situ* polymerization technique described in the text. The two vertical lines are adjacent cell walls which are separated by a distance of 1.12 mm.

Celcor would be a suitable stack material for a 10 kW thermoacoustic chiller (TRITON) that was under development at that time.

The measurements reported here separate the thermal and viscous effects by placing the Celcor sample at the center of a plane-wave resonator and examining the quality factor Q of the even and odd longitudinal modes separately. At low longitudinal mode numbers, the effects of thermal relaxation dominate the even modes when the sample is centered at a velocity node. The odd modes place the sample near a pressure node thereby enhancing the viscous loss mechanism and suppressing the thermal relaxation effects. This separation of thermo-viscous effects is only applicable when the length of the testing sample is much smaller than the wavelength. In our measurements, at the first mode ($n=1$), the length of the sample is 2.5% of the wavelength.

The measurement of quality factor, Q, was conducted at very low volume velocity. To explore the minor loss (end effect),[10] the driving voltage was increased from 100 mV to 3 V and the change of Q is less than 0.7% over that range of amplitudes. The minor loss is not a concern at the amplitudes used to conduct this experiment. The cutoff frequency above which nonplanar modes can propagate in the tube is 5900 Hz.

In this paper, we also investigate several pore sealing techniques, including different kinds of polymer solutions. We discover that even with the use of vacuum impregnation, the sealing of the ceramic pores is difficult to accomplish with any acceptable degree of uniformity. For that reason, we employed an *in situ* polymerization technique.

II. EXPERIMENTAL TECHNIQUE

A. Sealing

Sealing off the wall pores uniformly is important in comparing the measured data of coated sample with the porous sample. Shellac and several polymers: polystyrene $[\text{-}CH_2CH(C_6H_5)\text{-}]_n$, polymethyl methacrylate $[CH_2C(CH_3)(CO_2CH_3)\text{-}]_n$, polylethylene oxide $[\text{-}CH_2CH_2O\text{-}]_n$, and branched polyethylene $[\text{-}CH_2\text{-}CH_2\text{-}]_n$ were tried. Those pore sealants were applied using the following method: The Celcor samples are soaked in the polymer solution for hours then drained to keep the channels clear before drying it. But the wall pores could not be sealed uniformly inside the channels, even if the outside surface was coated with a uniform thin layer of the polymer. Vacuum impregnation and ultrasonic agitation were added to that procedure before draining the sample, but no obvious improvement was observed.

An *in situ* polymerization was found to work well. The Celcor sample was soaked in the solution of hydroxyethyl methacrylate azobisisobutyronitrile (AIBN). After the sample was drained, it was heated to 70 °C. At this temperature, the monomers linked together to form polymers with the help of initiator AIBN. Under microscopic inspection, the Celcor sample was coated uniformly both inside and outside. The micrographs of cell walls from the interior of test sample are shown in Fig. 2.

B. Quality factor of resonator

The quality factor, Q, of the resonant cavity is a dimensionless measure of the sharpness of the peak of the power curve of a resonance when amplitude is plotted versus frequency. There are many equivalent ways of expressing the Q. This multiplicity allows us to connect the most convenient experimental method to the parameter of interest. If a resonance is excited and then allowed to freely decay, Q is the ratio of 2π times energy stored to energy dissipated per cycle. For a cylindrical resonator that is long enough to neglect end effects, Q is proportional to R/δ. R is the radius of the resonator and δ is the penetration depth.[11] The penetration depth is inversely proportion to $\sqrt{n}$ for all the plane wave resonances where n is the longitudinal mode number. For all the plots in this paper, $Q/\sqrt{n}$ will be plotted versus mode numbers.

C. Instrumentation and modal analysis

The experimental apparatus is shown schematically in the Fig. 3. A 0.7-m-long cylindrical resonator with 0.033 m inner diameter was used.[12] A loudspeaker that was modified by Fitzpatrick[13] drove one end of the resonator and the re-

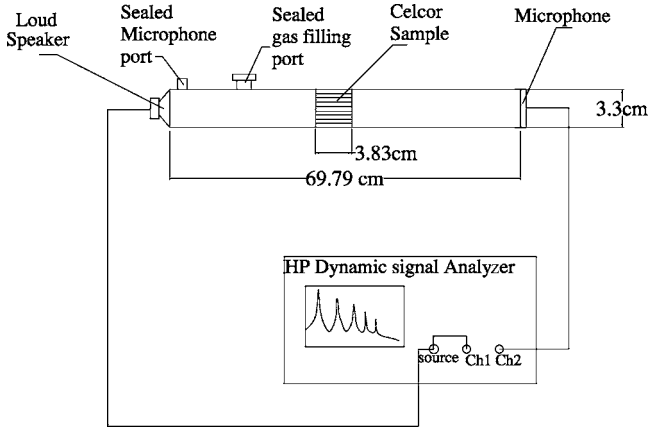


FIG. 3. Schematic diagram of the resonator, transducer, and spectrum analyzer used to make measurements on a coated and uncoated Celcor sample. Resonator dimensions are not drawn to scale.

sponse was measured by an electret microphone that was charged by liquid contact[14] located at the other end of the resonator. The microphone had the same cross-sectional area as the tube. An HP 3562A Dynamic Signal Analyzer was used to measure the transfer function between speaker input voltage and microphone output. A nonlinear pole-zero curve fit[15] was used to determine two complex poles that are complex conjugates, $a \pm ib$. In terms of these poles, the resonance frequency can be expressed as $f=\sqrt{a^2+b^2}$ and the quality factor as $Q=(-1/2a)\sqrt{a^2+b^2}$.

Placing a Celcor sample at the center of the cylindrical resonator, per the technique described by Moldover *et al.*,[16] the thermoviscous losses due to the Celcor sample could be evaluated by comparing the Q of the resonator with Celcor sample to the Q of the empty resonator. The theoretical model for resonance frequency and quality factor was developed using the DELTAE software.[17]

The pressure and velocity profile of the first two modes of the resonator is shown in Fig. 4 as well as an approximation by a lumped element model. For all of the odd modes, the Celcor sample is located at the velocity antinode but pressure node. This corresponds to the maximum viscous loss but minimum thermal relaxation loss. For all the even modes, the Celcor sample is located at the velocity node but pressure antinode to isolate the loss mechanism of interest in this study: the change in thermal conduction (Kirchoff) loss in porous or impenetrable walls.

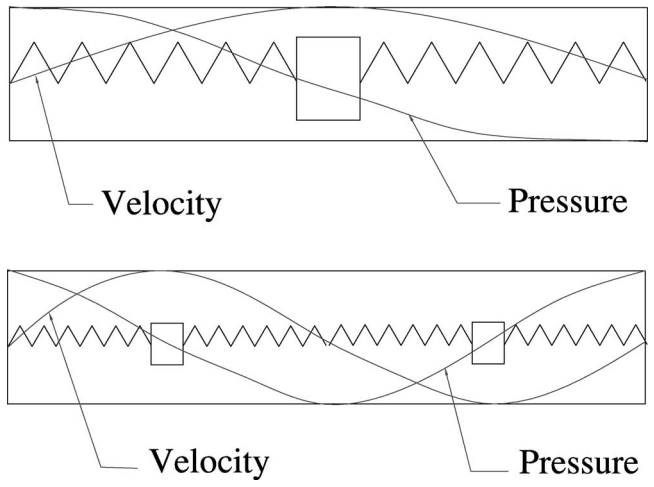


FIG. 4. Mode shape and lumped element (mass-spring) representation of the first two plane wave modes of the resonator with perfectly rigid boundary conditions at both ends.

D. The DELTAE model

The theoretical results are calculated using DELTAE. DELTAE is a partial differential equation solver which solves the one-dimensional wave equation based on the usual low-amplitude acoustic approximation and takes the thermal loss and viscous loss into consideration. It solves the wave equation in a gas or liquid, in a geometry given by the user, specified as a sequence of segments, such as ducts, compliances, transducers, and thermoacoustic stacks or regenerators. DELTAE does not include any nonlinear effects that arise at high amplitudes. As indicated previously in conjunction with the absence of "minor loss" effects, the system behavior was observed to be linear.

A five-segment DELTAE model was developed for the empty resonator and a seven-segment model for the resonator with a Celcor sample at the center. The Celcor sample was modeled as a STKREC segment using the pore geometry information provided by the manufacturer. A custom thermophysical parameter file was created based on the ceramic's heat capacity at 25 °C (c_p=616.2 J/K kg) and thermal conductivity (κ_s=2.6 W/K m) as provided by the manufacturer.[18] A solution to the appropriate *1*D wave equation is found for each segment, with complex pressures and volume flow rates matched at the junctions between segments.

The quality factor, Q, was calculated from the DELTAE output using the rate-of-change of phase with respect to frequency, calculated at the resonance frequency,

$$Q = \frac{f_n}{2} \cdot \left. \frac{d\phi}{df} \right|_{f_0}. \tag{1}$$

f_n is the resonance frequency of the nth mode and ϕ is the phase in radians between pressure and the velocity at the driving point.

III. RESULTS

The test results for the resonance frequency and quality factor for the seven lowest longitudinal modes of the empty resonator, the resonator with uncoated sample, and the resonator with coated sample were summarized in Table I.

Figure 5 compares the DELTAE prediction with the measured quality factor for the seven lowest frequency plane wave modes. As expected, the Q predicted by DELTAE exceeds the measured values. We attribute the majority of this discrepancy to the existence of various small gaps between the two 1/2 in. microphone ports, a 1/4 in. gas fill port, and the plugs used to seal those ports. Those gaps provide additional surface area for both thermal and viscous surface losses. We are convinced that the gaps between the resonator end caps, and the possibility of flow into small volumes created by spaces between the o-ring grove and o-ring were not responsible. Those seals were opened and closed several

TABLE I. Summary of the measured resonance frequency and the quality factor (Q) of the first seven modes for the empty resonator, the resonator with uncoated sample, and the resonator with coated sample.

Mode	Empty resonator		Resonator with uncoated sample		Resonator with coated sample	
	f(Hz)	Q	f(Hz)	Q	f(Hz)	Q
1	238.4	52.1	230.7	17.2	231	16.9
2	477.6	81.4	479.5	55.5	483.7	55.8
3	717	99.4	699.5	34.2	700.5	34.2
4	956.7	110.9	960.6	70.9	968.8	72.6
5	1197	120.8	1173	45.5	1174.8	45.3
6	1473	126.1	1440	77.3	1453.4	79.7
7	1678	134.7	1648.8	56.9	1654.8	56.3

times in the course of these experiments to change samples and no change in the empty resonator Q was ever observed. Due to this reproducibility, we felt it was appropriate to introduce an excess loss that would account for nonideal resonator behavior.

The plugging of the microphone ports and the gas fill port were not disturbed during all the measurements. The DELTAE model correctly predicts the energy stored per cycle in the resonator as evidenced by a ±0.03% agreement between the measured resonance frequency of the empty resonator and the DELTAE model frequencies. Only the dissipated power due to the small gaps at the gas filling port and microphone ports were underestimated because of the difficulty to include the accurate details of these small gaps in the model. The discrepancy between DELTAE prediction and measurement was calculated as Q_{Excess} at each resonance frequency as shown in Eq. (2). Q_{Excess} is used to correct the DELTAE prediction in cases which included the Celcor samples,

$$\frac{1}{Q_{\text{Measured}}} = \frac{1}{Q_{\text{DELTAE}}} + \frac{1}{Q_{\text{Excess}}}. \tag{2}$$

A. Resonator frequency shifts with Celcor sample

As shown in Fig. 4, for the odd modes, the center of the sample is located at a pressure node and velocity antinode. At all the even modes, the center of the sample is located at the pressure antinode but the velocity node. When the Celcor sample was placed at the center of the resonator, both the resonance frequency and the quality factor of the resonator were changed due to the volume exclusion and the excess losses introduced by the Celcor sample.

The change of the resonance frequency compared with the empty resonator is plotted using the square symbols in Fig. 6. After the Celcor sample was inserted, the resonance frequencies for all odd modes were decreased and the resonance frequencies for all even modes were increased. When the Celcor sample is located at the velocity antinode, the decrease of the frequency of the odd modes was produced by the smaller cross-section of the channel. This increases the "lumped inertance" based on the model of Fig. 4. The increase of the resonance frequency at all even modes was due to the volume exclusion produced by the sample at the pressure antinode, making the "lumped compliance" of the gas smaller in that region. We use this lumped parameter analysis to motivate our interpretation of these resonance frequency shift effects introduced by the presence of the Celcor sample. The DELTAE model produces a complete one-dimensional standing-wave solution that matches the boundary condition at both ends of the resonator as well as preserving continuity of complex pressure and volume velocity within both the

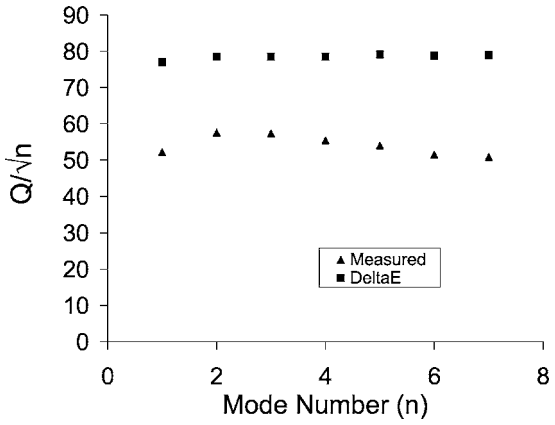


FIG. 5. Quality factors divided by the square-root of mode number (n) for the empty resonator as measured and calculated by DELTAE.

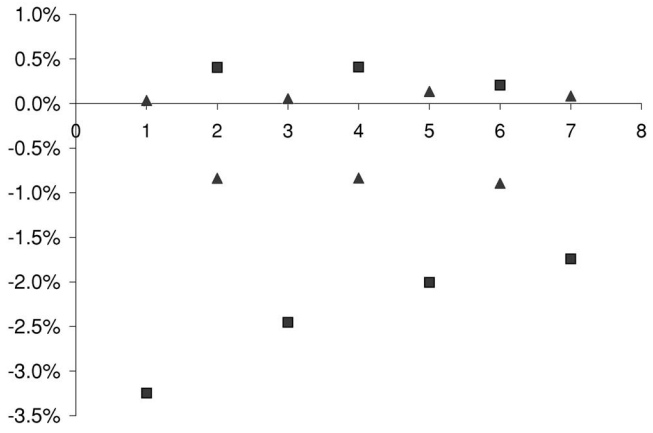


FIG. 6. Squares: the change in resonance frequency relative to the resonance frequency of the empty resonator, $(f_{\text{uncoated}}-f_{\text{empty}})/f_{\text{empty}}$, after the Celcor sample was inserted. Triangles: the frequency difference relative to DELTAE prediction for the resonator with uncoated Celcor sample $(f_{\text{uncoated}}-f_{\text{DELTAE}})/f_{\text{DELTAE}}$.

empty resonator and resonator with the Celcor inserted at the center.

The triangle symbols in Fig. 6 showed the frequency difference relative to the DELTAE prediction for the resonator with uncoated Celcor sample. It shows that the discrepancies at the even modes, when the sample is located at the pressure antinode, are larger then the discrepancies at the odd modes. This frequency depression for the even modes is expected, since the STKRECT segment in the DELTAE model assumes that the walls are rigid and impervious. The consistent 0.8% depression of the even-mode frequencies suggests that wall porosity can be inferred from this acoustic measurement. Modifying the total porosity used in the DELTAE model to match the measured resonance frequencies at even modes, the total porosity inferred was $\Omega_T = 0.89 \pm 0.02$. This agrees with the value $\Omega_T = 0.82$ determined by Arnott *et al.*, and $\Omega_T = 0.87 \pm 0.05$ determined by Champoux and Stinson.[19] After the wall pores were sealed off, the discrepancy in the even mode frequencies was entirely eliminated, as shown in Fig. 8.

After the Celcor sample was inserted in the resonator, the quality factors at each resonance frequency were expected to decrease. The quality factor Q of the empty resonator is

$$\frac{1}{Q} = \frac{\delta_v}{R} + \frac{\delta_\kappa}{R}\frac{(\gamma-1)}{(1+\epsilon_s)} + \frac{2\delta_\kappa}{L}\frac{(\gamma-1)}{(1+\epsilon_s)}, \tag{3}$$

where L is the length of the resonator, δ_v is the viscous penetration depth, δ_κ is the thermal penetration depth, and γ is the ratio of specific heats. The ratio of the thermally active heat capacity of the gas to that of the solid is given by ϵ_s which indicates how well the solid can keep the gas at the interface isothermal.

$$\epsilon_s = \frac{\rho_m c_p \delta_\kappa}{\rho_s c_s \delta_s}\frac{\tanh[(1+i)r_h/\delta_\kappa]}{\tanh[(1+i)l/\delta_s]}, \tag{4}$$

the subscript s stands for the properties of the solid and l is the half-thickness of the solid walls. This effect is included in the DELTAE model but, in this case, is negligible: $\epsilon_s \cong 0.003$.

For the empty resonator, since Q is inversely proportional to the penetration depth, the Q is proportional to $\sqrt{f}$. The frequency shift due to the insertion of the Celcor sample is shown by the square symbols in Fig. 6. Before comparing the empty resonator to the resonator with sample, the quality factor of the empty resonator was modified by $dQ/Q = 1/2\,(df/f)$. This modification of Q due to the slight change of frequency is also applied for the comparison between uncoated sample and DELTAE prediction.

The square symbols in Fig. 7 show the relative decrease of the measured quality factors, $(Q_{\text{uncoated}} - Q_{\text{empty}})/Q_{\text{empty}}$, after the sample was inserted compared to the empty resonator. The viscous (Stokes) losses produced by the Celcor is more significant than the thermal relaxation (Kirchoff) loss. There is also a trend apparent in the square symbols in Fig. 6 and other plots. The trend shows that the difference in behavior of the even modes and the odd modes becomes less obvious at higher frequencies. This is due to the length of the Celcor sample, which is becoming comparable to the wavelength at higher frequencies. For example, at the first mode, the length of the sample is about 2.5% of the wavelength. At the seventh mode, the length of the sample is about 18% of the wavelength.

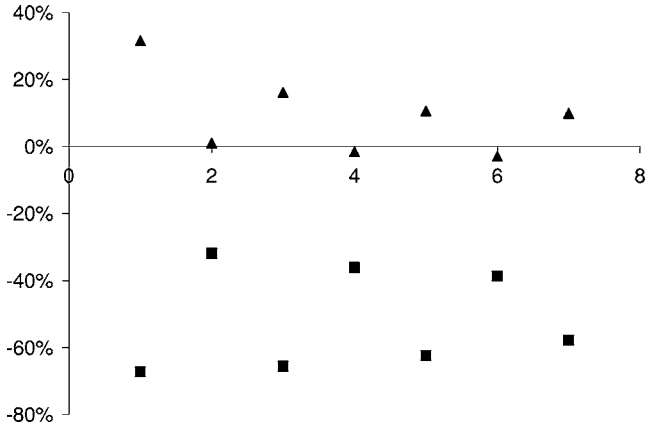


FIG. 7. Squares: relative change in measured quality factor compared to the empty resonator after the uncoated Celcor sample was inserted in the resonator, $(Q_{\text{uncoated}} - Q_{\text{empty}})/Q_{\text{empty}}$. Triangles: quality factor of the resonator containing the uncoated Celcor sample compared to the DELTAE prediction $(Q_{\text{uncoated}} - Q_{\text{DELTAE}})/Q_{\text{DELTAE}}$.

The measured quality factor exhibits fairly good agreement with the DELTAE prediction for all the even modes after correction for excess loss in the empty resonator. This is shown in the triangle symbols in Fig. 7.

B. The coated Celcor sample

For an uncoated Celcor sample, the favorable agreement of Q between the measured data and DELTAE model for all the even modes, along with the larger discrepancy of the resonance frequency at all the even modes, encourages a further exploration of the wall pores effects using this technique.

After coating the same Celcor sample with a very thin polymer layer to block only the wall pores, the quality factor and the resonance frequencies were measured again in the same resonator. The results are shown in Figs. 8 and 9.

Figure 8 shows the frequency changes compared with

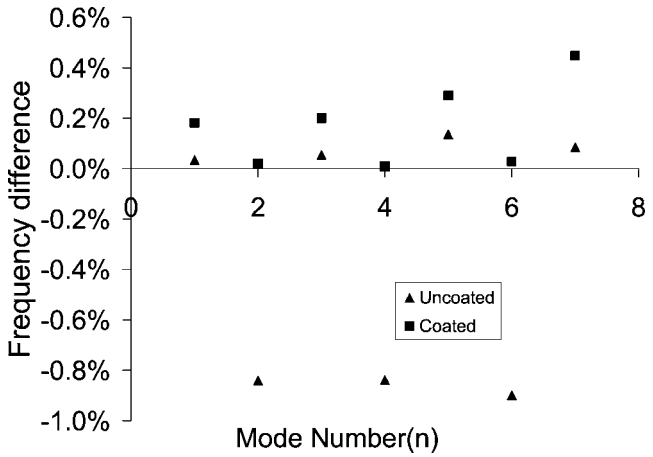


FIG. 8. Comparison of the relative frequency shift of coated and uncoated sample to DELTAE prediction.

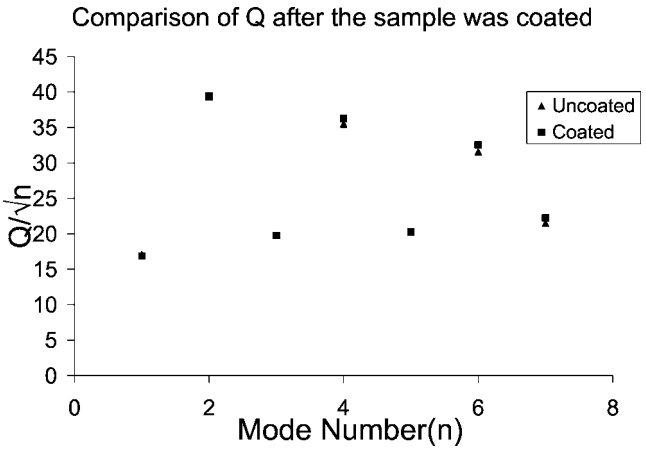


FIG. 9. Quality factor for the resonator with the coated and uncoated Celcor sample.

DELTAE prediction for both coated and uncoated samples. The discrepancies at all the even modes decreased to less than 0.1% after the sample was coated. Blocking the wall pores makes the sample "stiffer" thus increasing the resonance frequency to the value expected for a sample with "rigid, impenetrable" walls.

Figure 9 compares the measured Q for the uncoated sample to coated Celcor samples. At the lowest frequency, which is 230 Hz, there is only a 0.49% change for thermal relaxation losses. This is the measurement we sought in our effort to determine whether the Celcor stack that was to be used in the TRITON thermoacoustic air conditioner should be coated to reduce loss. These results assure us that coating of the Celcor ceramic was unnecessary to reduce loss in a standing-wave, thermoacoustic refrigerator application.

IV. CONCLUSION

The goal of this investigation was to determine if wall pores would introduce extra losses for the Celcor stack that was to be used in a thermoacoustic air conditioner. To do this, the thermal loss and the viscous loss were measured separately for both coated and uncoated Celcor sample. Wall pores were blocked by an *in situ* polymerization sealing process. Thermoviscous losses were also computed with DELTAE and compared to measured losses. There was no observed decrease in attenuation of even modes (dominated by thermal loss) for the coated sample below 1.0 kHz. The volume exclusion due to pore sealing was inferred by the increase in even mode resonance frequencies and was found to be in good agreement with other determinations of wall porosity.

ACKNOWLEDGMENTS

Professor Richard Raspet provided very useful comments regarding the relation of this work to the previous measurements made at NCPA. We are grateful for the support of this investigation provided by the Office of Naval Research, the Educational and Foundational Research Program funded by the Penn State Applied Research Laboratory, and an endowment provided by the United Technologies Corporation.

[1]S. L. Garrett and R. -L. Chen, "Build an 'Acoustic Laser'," Echoes **10**, 4–5 (2000); "Solar/heat driven thermoacoustic engine," *Proceedings of the 16th International Congress on Acoustics and 135th Meeting of the Acoustics Society of America*, Vol. **II**, Seattle, WA, 20–26 June 1998, pp. 813–814.

[2]R. A. Johnson, S. L. Garrett, and R. M. Keolian, "Thermoacoustic cooling for surface combatants," Nav. Eng. J. **112**, 335–345 (July 2000); S. F. Christie and J. C. Nilsen, **112**, 94–99 (Sept. 2000).

[3]W. V. Slayton and J. C. H. Zeegers, "An aeroacoustically driven thermoacoustic heat pump," J. Acoust. Soc. Am. **117**, 3628–3635 (2005).

[4]H. S. Roh, W. P. Arnott, J. M. Sabatier, and R. Raspet, "Measurement and calculation of acoustic propagation constants in arrays of small air-filled rectangular tubes," J. Acoust. Soc. Am. **89**, 2617–2624 (1991).

[5]W. P. Arnott, J. M. Sabatier, and R. Raspet, "Sound propagation in capillary-tube-type porous media with small pores in the capillary walls," J. Acoust. Soc. Am. **90**, 3299–3306 (1991).

[6]M. Bernard, D. Velea, and J. M. Sabatier, "Permanent removal of the wall porosity in monolithic catalyst support ceramics," J. Acoust. Soc. Am. **99**, 2430–2432 (1996).

[7]S. L. Garrett, "Resource Letter: TA-1: Thermoacoustic engines and refrigerators," Am. J. Phys. **72**, 11–17 (2004). The difference between stack-based (standing wave) and regenerator-based (traveling wave) thermoacoustic devices is addressed in the introduction.

[8]S. L. Garrett, J. A. Adeff, and T. J. Hofler, "Thermoacoustic refrigerator for space applications," J. Thermophys. Heat Transfer **7**, 595–599 (1993).

[9]S. L. Garrett, "High power thermoacoustic refrigerator," U.S. Patent No. 5,647,216 (1997); South African Letters Patent No. 96/6512 (1998).

[10]G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, New York, 2002); ISBN 0-7354-0065-2.

[11]G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145–1180 (1988).

[12]C. L. Burmaster, "Reciprocity calibration in a plane wave resonator," Ph.D. dissertation in Engineering Acoustics, US Naval Postgraduate School, Monterey, CA, December 1985. Also available as DTIC Report No. AD A164 149.

[13]M. Fitzpatrick, "Electrodynamic driver for the Space Thermoacoustic Refrigerator (STAR)," Master's thesis in Physics, US Naval Postgraduate School, Monterey, CA, March 1988. Also available as DTIC Report No. AD A192 337.

[14]P. W. Chudleigh, "Mechanism of charge transfer to a polymer surface by a conducting liquid contact," J. Appl. Phys. **47**, 4475–4483 (1976).

[15]"HP 35665A Dynamic Signal Analyzer Concepts Guide," Hewlett-Packard, 8600 Soper Hill Road, Everett, WA 98205-1298, p. 15-9.

[16]M. R. Moldover, K. A. Gillis, J. J. Hurley, J. B. Mehl, and J. Wilhelm, "Acoustic measurements in gases," in *Modern Acoustical Techniques for the Measurement of Mechanical Properties*, edited by M. Levy, H. E. Bass, and R. Stern (Academic, San Diego, 2001), Chap. 10, pp. 395–399.

[17]The current version (5.1b2) of the DELTAE software and manual are available on the CD-ROM that is included in the textbook by G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, New York, 2002); ISBN 0-7354-0065-2. The software can also be downloaded from the Los Alamos National Laboratory website: www.lanl.gov/thermoacoustics/DeltaE.html.

[18]Celcor, Corning Environmental Products Div., Corning, Inc., HB-CB-03-1, NY 14831.

[19]Y. Champoux, M. R. Stinson, and G. A. Daigle, "Air-based system for the measurement of porosity," J. Acoust. Soc. Am. **89**, 910–916 (1991).

Acoustic measurement of lubricant-film thickness distribution in ball bearings

Jie Zhang and Bruce W. Drinkwater
Department of Mechanical Engineering, University Walk, University of Bristol, Bristol BS8 1TR, United Kingdom

Rob S. Dwyer-Joyce
Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, United Kingdom

(Received 6 July 2005; revised 28 October 2005; accepted 9 November 2005)

An oil-film thickness monitoring system capable of providing an early warning of lubrication failure in rolling element bearings has been developed. The system is used to measure the lubricant-film thickness in a conventional deep groove ball bearing (shaft diameter 80 mm, ball diameter 12.7 mm). The measurement system comprises a 50 MHz broadband ultrasonic focused transducer mounted on the static outer raceway of the bearing. Typically the lubricant-films in rolling element bearings are between 0.1–1.0 μm in thickness and so are significantly smaller than the ultrasonic wavelength. A quasistatic spring model is used to calculate oil-film thickness from the measured reflection coefficient data. An accurate triggering system has been developed to enable multiple reflection coefficient measurements to be made as the contact ellipse sweeps over the measurement location. Experiments are described in which the loading conditions and rotational speed are varied. Lubricant-film thickness distributions measured ultrasonically are described and are shown to agree well with the predictions from classical elastohydrodynamic (EHD) lubrication theory, particularly at high radial loads and low rotary speeds. A range of parameters affecting the performance of the measurement are discussed and the limits of operation of the measurement technique defined. [DOI: 10.1121/1.2146109]

PACS number(s): 43.35.Yb, 43.35.Zc, 43.35.Ns [TDM] Pages: 863–871

I. INTRODUCTION

Fluid film lubrication occurs when opposing bearing surfaces are completely separated by a lubricant film, typically an oil or grease, which reduces friction and wear and provides smooth running and a satisfactory life for machine elements. The consequences of either film failure or the absence of a film are usually manifested by severe friction and wear[1] and the eventual failure of the bearing. Monitoring of the condition of bearings is therefore of major interest to a range of industries, particularly those in the aerospace, marine, power generation, and process sectors. In the industry-bearing condition, monitoring is currently achieved by the measurement of running temperature, vibration signals, or acoustic emission. The disadvantage of all these techniques is that they are measuring the effect of an already partially failed bearing. A typical failure scenario in the event of a loss of lubricant is an elevated temperature caused by contact of the bearing surfaces, followed by a collapse of the oil film, surface spalling or pitting, increased vibration, and possible acoustic emission.

Measurement of the lubricant-film thickness provides a more direct and quantitative way of monitoring the performance of a bearing, before actual bearing damage occurs. However, this is a challenging measurement as typically the load is carried by an extremely thin oil film over a small lubricated region. For example, in a 6016 ball bearing (shaft diameter 80 mm, ball diameter 12.7 mm), the contact ellipse is 0.3×3 mm under a typical (15 kN) operating load, and the oil film thickness is in the range of 0.1–1.0 μm.[2] In this range, either electromagnetic or optical methods have been used to measure the lubricant film thickness. However, these techniques suffer from serious drawbacks. The resistance[3] and capacitance methods[4,5] require either an insulated surface mounted sensor or complete electrical isolation of the contact elements. These methods are generally limited to lubricant films above about 1 μm in thickness.[6] Optical interferometry[7] and optical fluoresence techniques[8] have also been used in test bearings, but these require a transparent window through which to make the measurement. These requirements mean that both electromagnetic and optical methods are rarely used outside the laboratory.

Various researchers have used reflection coefficient measurements to infer the properties of thin layers, adhesive layers, surface coatings, and adhesive interlayers.[9–12] Recently, a number of workers have used similar approaches to characterize lubricant films in a number of different machine elements. Anderson *et al.*[13] used the transmission and reflection of ultrasonic waves to monitor the collapse of oil films in thin shaft seals. Dwyer-Joyce *et al.*[14] and Zhang *et al.*[15] used ultrasonic reflection coefficient measurements to monitor the lubricant-film thickness in thick film bearings such as journal bearings and thrust-pad bearings. In these bearings the lubricant film was in the range 1.0–20 μm and the film was constant with respect to the fixed transducer. The results were shown to agree well with models of the bearing performance. Dwyer-Joyce *et al.* did attempt to measure the lubri-

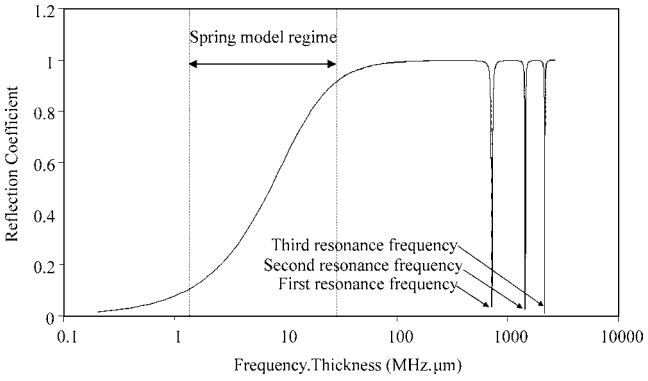


FIG. 1. Predicted reflection coefficient spectrum for a layer of mineral oil between two steel half-spaces.

cant thickness in ball bearings, however, success was limited as no satisfactory triggering arrangement was developed. This resulted in point measurements of oil-film thickness with little or no control over the measurement location. The rolling element bearing application was found to be considerably more challenging than the previously published hydrodynamic applications, as the contact patches are moving relative to the fixed transducer and the size of the contact and thickness of the lubricant in the contact are an order of magnitude lower.

In this paper we take a lubricant-film thickness measurement in rolling element bearings a critical step forward and show for the first time that ultrasound can be used to measure the central film thickness and the film thickness distribution robustly and accurately.

II. BACKGROUND THEORY

A. Ultrasonic reflection from an oil film

The structure of a rolling element bearing can be represented as a multilayered system consisting of an "outer raceway"–"lubricant"–"ball"–"lubricant"–"inner raceway." When an ultrasonic pulse propagates through this structure, ultrasound will be reflected from all the interfaces within the system, including from the lubricant layer. If the thickness of the lubricant layer is small in comparison with the ultrasonic wavelength, it can be shown that the reflection of ultrasound is governed only by its stiffness (see, for example, Ref. 16). The normal stiffness of a fluid layer, K_N, is given by[14]

$$K_N = \frac{B}{h}, \tag{1}$$

where h is the lubricant-layer thickness, B is the bulk modulus of the fluid (and $B=\rho c^2$), and ρ and c are the density and longitudinal wave speed of the fluid, respectively. Assuming that the media on either side of the layer have identical acoustic properties and that the wave is normally incident, the lubricant-film thickness can be extracted from the well known quasistatic spring model[15,16] as

TABLE I. Acoustic properties of Shell T68 lubricating oil and steel.

	Density ρ (kg/m^3)	Longitudinal wave velocity c (m/s)	Bulk modulus B (GPa)
Oil at 0.1 MPa	876	1460	1.84
Oil at 0.8 GPa	1002	3550	12.6
Oil at 1.5 GPa	1044	4500	21.2
Steel (EN24)	7900	5900	172

$$h = \frac{B}{\pi f z}\sqrt{\frac{|R(f)|^2}{1-|R(f)|^2}}, \tag{2}$$

where z is the acoustic impedance of the media surrounding the lubricant film, $R(f)$ is the amplitude of the measured reflection coefficient, which is a function of the ultrasonic frequency, f. In general, the reflection coefficient is measured by comparing the signal reflected from the interface of interest to that from a known reference interface:

$$R(f) = \frac{A_m(f)}{A_{\mathrm{ref}}(f)}R_{\mathrm{ref}}, \tag{3}$$

where $A_m(f)$ is the amplitude of the signal reflected from the lubricant-film layer, $A_{\mathrm{ref}}(f)$ is the amplitude of the reference signal, and R_{ref} is the reflection coefficient of the reference interface. The reflection coefficient calculated from Eq. (3) can then be used in Eq. (2) to extract the lubricant film thickness, assuming all other material constants, acoustic properties, and the reference reflection coefficient are known.

Figure 1 shows a plot of the steel–oil–steel reflection coefficient over a wide range of frequency–thickness products, where frequency–thickness is plotted to generalize the results. This was calculated using a three-layer continuum model[17] with the material property data shown in Table I. Figure 1 also indicates the region over which the quasistatic spring model can be applied. In their recent paper, Zhang *et al.*[15] analyzed the error inherent in the use of the quasistatic spring model to calculate the layer thickness from reflection coefficients. For example, they showed that if the reflection coefficient could be measured to 1% accuracy and was in the range $0.1<R<0.95$, then the thickness error was 2%. These limits are shown in Fig. 1 as the frequency–thickness product range, $1.3<hf<38$ MHz μm. For the 50 MHz transducer used in this paper (which had usable frequency bandwidth 30–75 MHz) and detailed in Table II this equates to a lubricant-film thickness range of 0.02–1.3 μm. It is the aim of this paper to demonstrate that such a measurement can be realized and that this can be performed on a rotating element bearing operating under typical conditions.

TABLE II. Transducer characteristics and dimensions.

Center frequency f_c (MHz)	Wavelength in water at F_c λ (μm)	Focal length F (mm)	Element radius of curvature D (mm)	Active element diameter d (mm)
50	30	23	25	5

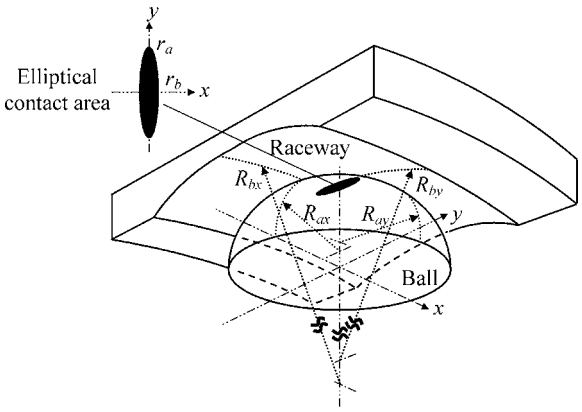


FIG. 2. Geometry of a ball bearing outer raceway and ball in contact.

B. Oil-film formation in a ball bearing

The geometry of the contact that forms when a ball is pressed onto a closely conforming raceway groove can be calculated from the applied load, P.[1] The contact area is elliptical in shape with the major (r_a) and minor (r_b) semicontact radii, given by

$$r_a = \left(\frac{6k^2\epsilon PR'}{\pi E'}\right)^{1/3}, \quad r_b = \left(\frac{6\epsilon PR'}{\pi kE'}\right)^{1/3}, \tag{4}$$

where k and ϵ are measures of the shape of the contact ellipse obtained from look-up tables.[1] E' is the reduced elastic modulus and R' is the reduced radius of curvature, given by

$$\frac{1}{E'} = \frac{1}{2}\left(\frac{1-\nu_a^2}{E_a} + \frac{1-\nu_b^2}{E_b}\right),$$
$$\frac{1}{R'} = \frac{1}{R_{ax}} + \frac{1}{R_{bx}} + \frac{1}{R_{ay}} + \frac{1}{R_{by}}, \tag{5}$$

where E refers to the elastic modulus, and ν refers to Poisson's ratio. As shown in Fig. 2, the subscripts a and b refer to the two rolling elements (i.e., the ball and the raceway), and R_x and R_y refer to the radii of curvature in the x and y directions, respectively.

The geometry of the ball and raceway is such that the major semicontact width r_a is around ten times the minor semicontact width r_b. The ball sweeps across the raceway in the x direction so ultrasonic measurements are made across the minor axis of this ellipse. The pressure distribution over the elliptical region is similar to the dry contact case and given by[1].9

$$p = p_0\sqrt{1-\left(\frac{x}{r_b}\right)^2 - \left(\frac{y}{r_a}\right)^2}, \tag{6}$$

where p_0 is the maximum contact pressure that occurs at the center of the ellipse. The load on the ball is not the same as that on the whole bearing, because several balls are in contact with the raceways at any instant. The load on the maximum loaded ball, P, directly opposite the point of application of the bearing load is given by[18]

$$P = \frac{5W}{n}, \tag{7}$$

where W is the radial load on the whole bearing and n is the number of the balls in the bearing. The mean and peak contact pressures are then given by

$$p_m = \frac{P}{\pi r_a r_b} \quad \text{and} \quad p_0 = \frac{3P}{2\pi r_a r_b}. \tag{8}$$

In this way, Eqs. (6), (7), and (8) give the pressure distribution over the measured ball. For a ball bearing, operating in the elastohydrodynamic lubrication regime, the lubricant-film thickness can be estimated from the numerically derived regression equations of Dowson and Higginson.[1,19] They showed that the central film thickness, h_c, can be expressed as

$$\frac{h_c}{R'} = 2.69\left(\frac{U\eta_0}{E'R'}\right)^{0.67}(\alpha E')^{0.53}\left(\frac{P}{E'R'^2}\right)^{-0.067} \times(1-0.61e^{-0.73k}), \tag{9}$$

where U is the mean surface speed, η_0 is the lubricant viscosity at the contact entry, α is the pressure–viscosity coefficient, k is an ellipticity parameter, and P is the load on the measured ball obtained from Eq. (7).

Contact pressures in the ball bearing contacts are very high, and this has the effect of increasing both the density and the bulk modulus of the lubricant. Jacobson and Vinet[20] developed a model for this bulk modulus variation. They give an equation of state to describe the behavior of the lubricant under pressure, p:

$$p = \frac{3B_0}{s^2}(1-s)e^{t(1-s)}, \tag{10}$$

and the bulk modulus under pressure is given by

$$B = \frac{B_0}{s^2}[2+(t-1)s-st^2]e^{t(1-s)}, \tag{11}$$

where B_0 is the bulk modulus at zero pressure, t is a lubricant specific parameter, and s is a function of the relative compression:

$$s = \sqrt[3]{\frac{\rho_o}{\rho_p}}, \tag{12}$$

where ρ_0 is the density at zero pressure and ρ_p at pressure p. The parameter t is determined empirically from tests on lubricants in high-pressure cells (pressures up to 2.2 GPa were reported in Ref. 20). Because of the experimental complexities, this data is scarce; in this paper, the data available for an oil of a similar generic type to that used in the experiments is used. Table I shows the bulk modulus determined at three contact pressures using this analysis.

III. BALL BEARING EXPERIMENTAL APPARATUS

Experimental apparatus capable of accurately measuring the ultrasonic reflection coefficient from a lubricant film in the 6016 bearing system is shown in Fig. 3(a). The bearing rig consisted of a rotating shaft of 80 mm diameter supported

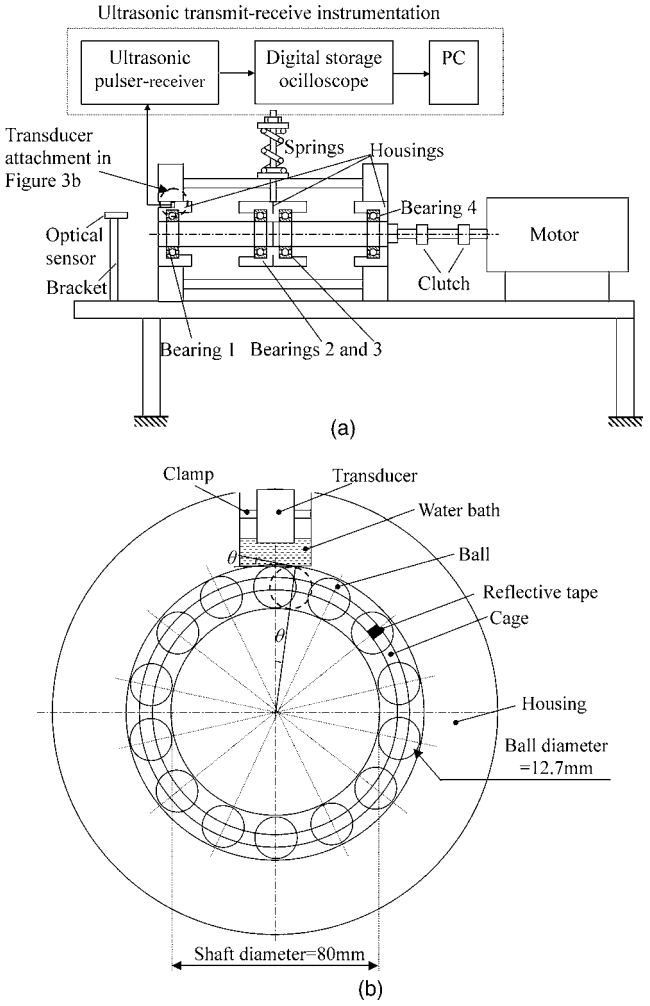


FIG. 3. (a) Schematic diagram of the experimental apparatus made up of four 6016 ball bearings. (b) Transducer attachment and bearing geometry.

on four 6016 ball bearings and ultrasonic transmit–receive instrumentation. As shown in Fig. 3(a), bearings 1 and 4 were fitted to the ends of the shaft and fixed into rigid housings. Radial loads were applied to the shaft through bearings 2 and 3. The load was applied vertically upward by an arrangement of springs. This meant that in bearings 1 and 4 the ball at the top of the raceway was the most heavily loaded. The rotary shaft speed was controllable in the range 100–2900 rpm by a 7.5 kW electric motor. The bearing was lubricated with Shell T68 mineral oil (details shown in Table I) via a total loss gravity feed system. An optical sensor was used, both to allow accurate triggering of the ultrasonic instrumentation and to measure the shaft speed. This was triggered by a reflective tape attached to the ball cage (which rotates at half the shaft speed). Bearing 1 was instrumented with the ultrasonic measurement system. Figure 3(b) shows the ultrasonic measurement system in more detail. A focused, longitudinal wave piezoelectric ultrasonic transducer was mounted in a circular hole in the housing such that it was normal to the top surface of the outer raceway. This mounting hole was 17 mm in diameter, compared to the housing width of 85 mm, and so the outer raceway of the bearing was well supported. This transducer, detailed in Table II, acted as both an emitter and receiver (pulse-echo mode), and had a center frequency of 50 MHz, an active element diameter of 5 mm, and a focal length in water of 23 mm. This equates to a theoretical focal spot size (defined at −6 dB down from the maximum) in the plane of the lubricant film of 146 μm at the center frequency.

This spot size defines the spatial resolution of the measurement system. The transducer was selected to be as high frequency as possible, while still able to operate in a regime where material attenuation was acceptable. The transducer was connected to an ultrasonic pulser-receiver (Panametrics 5072PR) that outputs a voltage spike containing energy in the range 5–200 MHz. This was used to excite the ultrasonic transducer, receive and amplify the reflected signals, which were then passed to a digital scope (sampling frequency 250 MHz) and PC for storage and analysis.

The reflective tape attached to the bearing cage is also shown in Fig. 3(b). When this tape passed the optical sensor it generated a 5 V positive trigger pulse. This pulse was used to trigger a signal generator (Agilent 33220A). After the addition of an adjustable delay, the signal generator then triggered the pulser-receiver at its maximum pulse repetition frequency, which was 20 kHz. By triggering in this way, a number of ultrasonic pulses were able to interrogate the lubricated "contact" region as it passed under the transducer. Because of memory limitations in the oscilloscope, only 80 reflected pulses could be stored in the oscilloscope at any time, and hence this is the maximum number of measurement points that can be obtained per ball. Hence, the actual number of points in a given measurement was governed by a combination of the pulser-receiver repetition rate and the speed at which the balls in bearing passed the transducer. The delay between the optical trigger and the triggering of the pulser-receiver was adjusted by knowledge of the shaft speed and the distance between the trigger point and the focal spot of the transducer.

In order to calculate the reflection coefficient, it is necessary to obtain a reference reflection from a known interface. The apparatus was used to obtain a steel-air reflection before a lubricant was introduced. The steel-air reflection coefficient is known to be 0.999 98. Using this reference all measured reflections were converted to reflection coefficient via Eq. (3), and this analysis was performed in the frequency domain.

It is well known that the mechanical and acoustic properties of mineral oils vary with both temperature and pressure. In this paper the measurements were taken over a short period of time (typically less than 30 min). Between measurements the system was allowed to cool to room temperature. During this time, the temperature of the bearing was measured and found to vary by 2 °C for the highest load and speed case. This temperature variation will have a small effect on the film thickness and acoustic properties and so this effect is neglected in the rest of this paper.

IV. RESULTS

A. Central thickness

Figure 4 shows the experimentally measured reflection coefficients for various shaft speeds from 106–506 rpm and

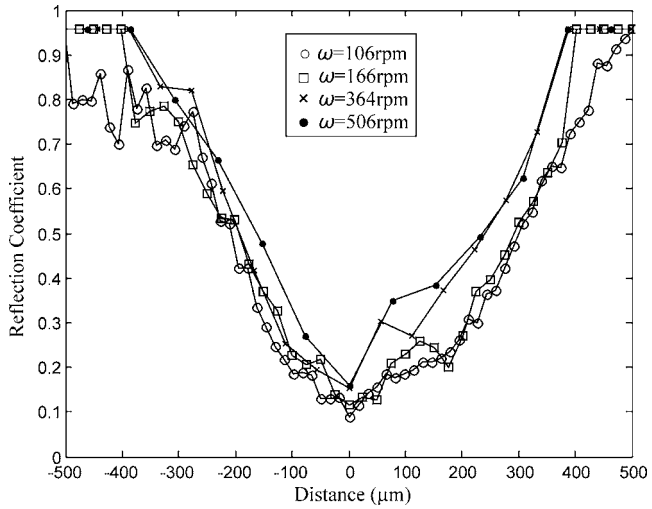


FIG. 4. Reflection coefficient recorded for various shaft speeds at a radial load of 15 kN.

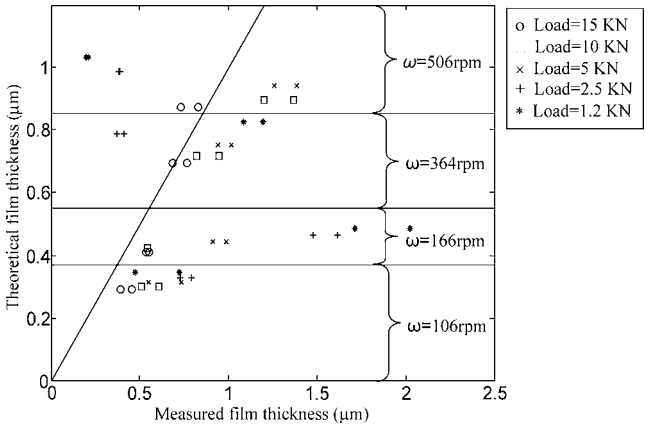


FIG. 5. A comparison of lubricant-film thicknesses measured by an ultrasonic means with EHD theoretical solution at various radial load and shaft speeds.

a test bearing radial load of 15 kN. Data is recorded as the ball passes under the measurement location, and so a reflection coefficient profile is created. It can be seen that the reflection coefficient falls to a minimum at the center of the lubricated contact region and then rises again.

The reflection coefficient recorded when the ball is immediately below the transducer (i.e., the center of the contact $x=0$ in Fig. 4) has been used to calculate the oil film thickness using Eq. (2). Figure 5 shows a comparison between this measured film thickness and the theoretical lubricant-film thickness [Eqs. (5), (7), and (9)] for a range of different operation conditions using the bearing data shown in Table III.

Different points for the same combination of speed and load represent separate readings and indicate the scatter in the measurement. For a given shaft speed, the experimental thickness decreases with increasing radial load. These experimentally measured trends can be seen to be in reasonable agreement with those predicted by the elastohydrodynamic theory. It is also worth noting that there is better quantitative agreement at high loads and low speeds. In particular, the results at 1.2 and 2.5 kN are poor in comparison to those at higher loads. Possible reasons for the discrepancy observed in Fig. 5 are explored in detail in Sec. V.

B. Thickness distribution

The measured reflection coefficient profiles shown in Fig. 4 are now used to calculate the oil film thickness distribution across the contact [again using Eq. (2)]. However, it is necessary to first account for the variation in the lubricant bulk modulus as it changes over the contact width. The pressure distribution across the contact width (in the x direction) is obtained from Eqs. (6), (7), and (8). It is assumed that the measured reflection results are recorded along the ellipse axis at $y=0$. The bulk modulus along this line is then determined using Eqs. (10), (11), and (12).

Figure 6(a) shows this bulk modulus distribution for several applied loads. It is worth noting that from Eq. (2), uncertainties in bulk modulus will translate directly into uncertainties in measure lubricant-film thickness.

Figure 6(b) shows lubricant-film thickness distribution extracted from the reflection coefficient profiles (shown in Fig. 4) using the bulk modulus distribution shown in Fig. 6(a). It can be seen that this thickness distribution is characterized by 10–80 measurement points, depending on the shaft speed. Because of the 20 kHz measurement interval, as the shaft speed increases, so the number of measurement points decreases. Also shown in Fig. 6(b) is the theoretical thickness distribution, assuming the central film thickness calculation [Eqs. (4), (5), (7), and (9)] is valid across the whole contact. Note that the bearing data shown in Table III was used in Eq. (9) to calculate the theoretical lubricant-layer thickness. It can be seen that there is good agreement at the center of the lubricated contact and that elsewhere the agreement is less good. In Sec. V we seek to explain the observed shape of this thickness distribution as well as discuss a range of measurement issues that may lead to uncertainties in the measured thickness.

V. DISCUSSION

In this section the acoustic pressure distribution over the focal spot is used to explain the shape of the experimentally

TABLE III. Parameters required to calculate the theoretical lubricant-film thickness via the Dowson and Higginson equation (Ref. 19).

Reduced modulus E' (GPa)	Reduced radius R' (m)	Pressure viscosity coefficient α (GPa^{-1})	Ellipticity parameter k	Simplified elliptical integrals ϵ	Effective viscosity η_0 (N/m^2 s)
228	5.85e-3	20	11.5	3.8	0.2

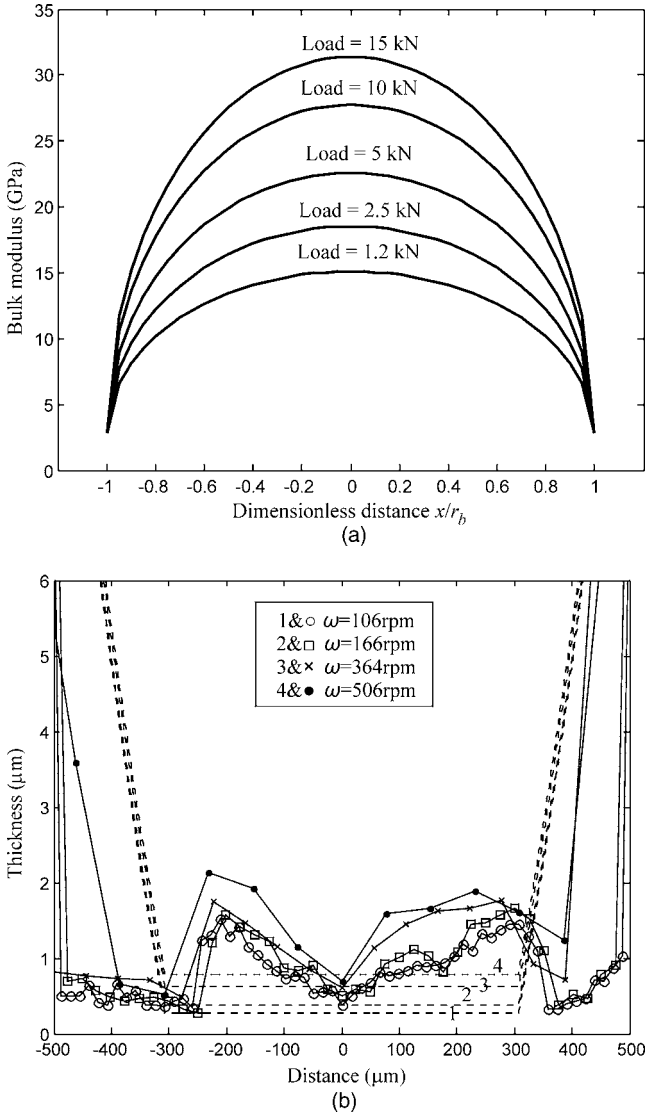


FIG. 6. (a) Bulk modulus distribution along the minor axis at various radial loads. (b) Measured oil film thickness recorded for various shaft speeds at a radial load of 15 kN. Results are compared with the theoretical thicknesses indicated by the dashed line. Note that in the key the numbers and symbols refer to theoretical and measured lines, respectively.

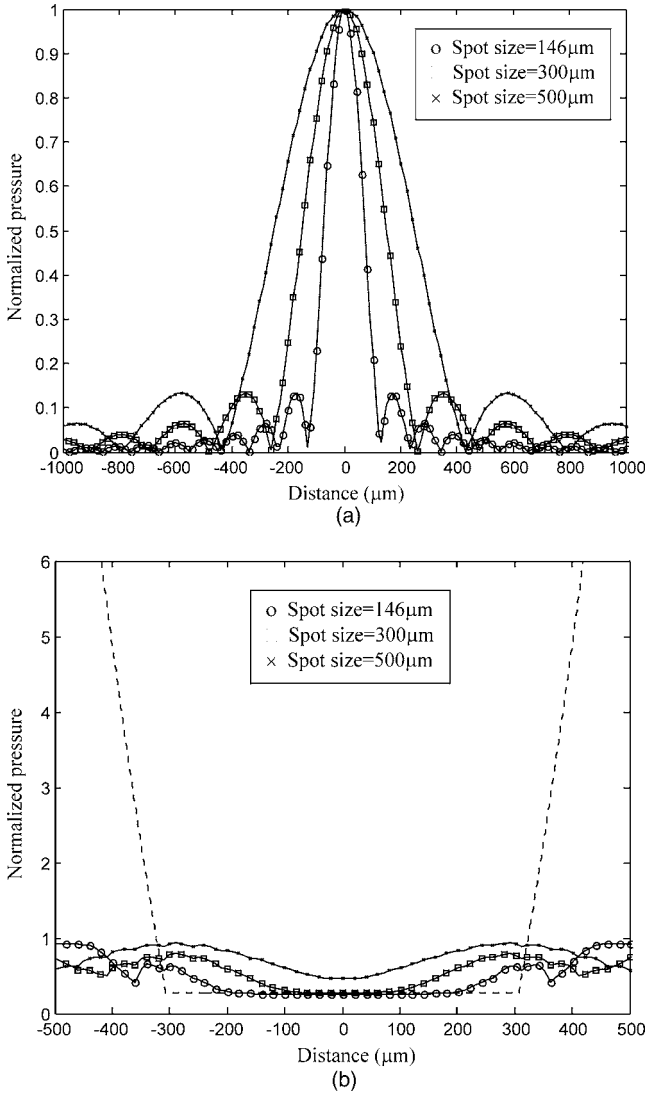


FIG. 7. (a) The acoustic pressure distribution at the focal plane for transducers with various spot sizes. (b) Lubricant-film thicknesses calculated by theory and using the weighting function of acoustic pressure distribution at the focal plane. Results from various spot sizes are compared.

measured lubricant-film distribution. Errors in the reflection coefficient due to the orientation of the transducer with respect to the lubricated contact are discussed and quantified. The limits of operation of the technique are then discussed.

A. Effect of the transducer focal zone size

When focused on the lubricant film, the ultrasonic transducer measures an average reflection coefficient, weighted by the acoustic pressure distribution in the plane of the lubricant film. The acoustic pressure distribution in the focal plane of a circular transducer is given by[21]

$$p(x) = \left| 2p_0 \left(\frac{\pi D x}{\lambda F} \right)^{-1} J_1 \left(\frac{\pi D x}{\lambda F} \right) \right|, \tag{13}$$

where p_0 is the center acoustic pressure (set to unity to normalize the pressure distribution), F is the focal length, x is the radial distance from the central axis, D is the diameter of the active element, λ is the wavelength, and J_1 is a first-order Bessel function. The application of Eq. (13) to the transducer used in this work, detailed in Table II, gives a theoretical focal spot width (measured at −6 dB of the maximum pressure) of 146 μm at 50 MHz. However, when this transducer was moved over a sharp edge, its focal spot size was estimated as being 500 μm. This discrepancy is probably due to a combination of a slight roundness present on the sharp edge and a small underperformance of the transducer. Figure 7(a) shows the acoustic pressure distribution predicted for the transducer data shown in Table II as well as for a transducer with 300 and 500 μm spot sizes.

The acoustic pressure distribution can now be used to weight the reflection coefficient predicted via the application of Eq. (2) to the theoretical lubricant-layer thickness [Eqs. (4), (5), (7), and (9)]. The weighting procedure takes the following form:

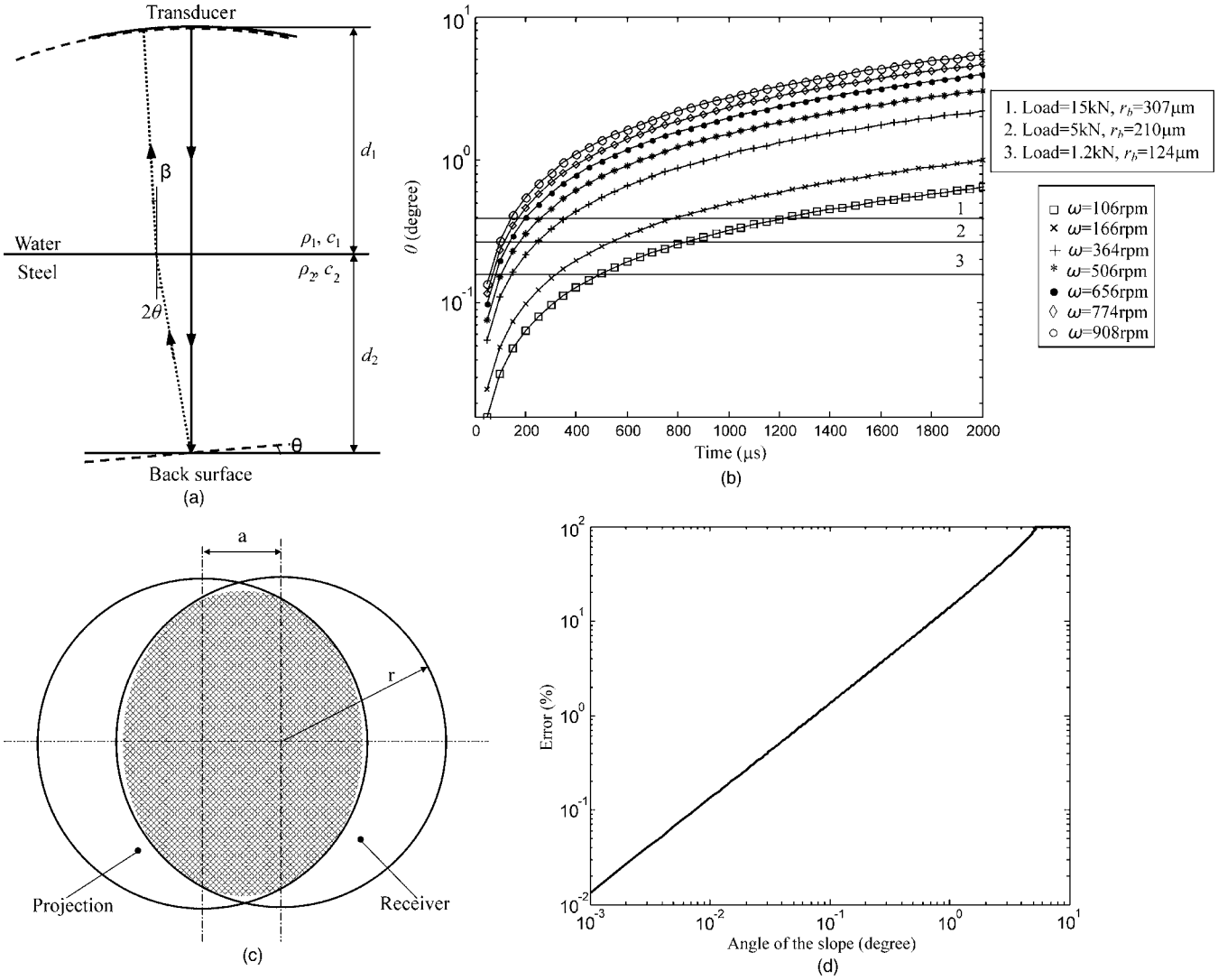


FIG. 8. (a) A comparison of the geometries of the reflected signal from a normal surface and a surface with a small slope. (b) The angle of the slope of lubricated contact area versus the time from TDC. The horizontal lines for various loads indicate the maximum angle in the lubricated contact area. (c) Schematic diagram of the geometries of the effective receiving area of the transducer. The cross-hatched area indicates the effective receiving area. (d) Reflection coefficient error caused by the surface slope of the lubricant contact area.

$$R_w(x_i) = \frac{\int_{-\infty}^{\infty} p^2(x - x_i)R(x)dx}{\int_{-\infty}^{\infty} p^2(x)dx}, \tag{14}$$

where R and R_w are the unweighted and weighted reflection coefficients, respectively, and x_i refers to x position at which the weighted average is calculated.

Figure 7(b) shows the simulated measured lubricant film thickness based on Eq. (9) and the weighted version using Eq. (14) and then Eq. (2). From Fig. 7(b) it can be seen that as the spot size increases, so the "curvature" of the thickness distribution increases. It is also apparent that the form of the thickness distribution predicted in Fig. 7(b) is similar to that measured experimentally and shown in Fig. 6(b). It can also be seen that the central value of lubricant-film thickness shows close agreement with the "theoretical" value as it is least affected by the focal spot size effect.

B. Effect of transducer orientation

As the ball bearing passes the measurement region the orientation of the lubricant film with respect to the transducer axis changes. The angle of this slope, θ, is shown in Figs. 8(a) and 3(b) and can be expressed as

$$\theta = \omega T = \omega N \Delta t, \tag{15}$$

where ω is the angular velocity of the ball and T is the time since the ball was at top dead center (TDC), i.e., directly below the central axis of the transducer. This can also be written in terms of the number of measurement points since TDC, N, and the sampling interval, Δt (50 μs). Figure 8(b) shows how this angle increases with increasing the time from TDC. Also shown in Fig. 8(b) are horizontal lines that show the maximum angle monitored within the contact patches generated at a number of different loads. The angular range from 0 (i.e., the contact patch is normal to the transducer axis) to these lines represents cases encoun-

tered in this paper. Ray tracing theory and angle-dependent spectral distortion (ASD) theory[22] can be used to analyze the effect of this slope on the measured reflection coefficient. There are two points in considering this slope effect, the projection of the ultrasound rays onto the transducer, and the active element size in the transducer. The projection angle φ can be found from

$$\varphi = \beta - \tan^{-1}\left(\frac{a}{F}\right), \tag{16}$$

where $a = d_2 \tan 2\theta + d_1 \tan \beta$, β is the angle of the central reflected ray in the water that can be obtained from Snell's law and the other parameters are defined in Fig. 8(a).

Figure 8(c) shows the geometries of the receiving zone of the transducer and the reflected ultrasonic beam projection. When a ball is at TDC, the receiving area and reflected projection overlap exactly. At a given projected angle, the reflected signal is reduced by

$$Q = 1 - \frac{2r^2 \cos^{-1}(a/2r) - \frac{a}{2}\sqrt{r^2 - a^2/4}}{\pi r^2} \cos \varphi, \tag{17}$$

where r is the radius of the active element. Figure 8(d) shows this as a relative error as a function of the angle of slope, θ. For example, if a lubricant film thickness measurement accuracy of 2% is required, the experimental data should be only captured when the angle of slope of the lubricated contact region is less than 0.2°. Note that this error will always occur at the edges of the lubricated contact region, where the angle of slope is greatest. Also note that this effect will tend to cause a reduction in the measured reflection coefficient and hence an underestimate of the lubricant film thickness. This is also good confirmation that the measured central thicknesses is the least prone to error and therefore most suitable for comparison with theoretical values.

C. Limits of operation

At present, the measurement system is limited to a time interval of 50 μs by the maximum pulse repetition frequency (20 kHz) achievable by the pulser receiver. If this could be improved, then the next limit would be due to the ultrasonic waves reverberating in the coupling water and the outer raceway and hence not decaying to zero before the next measurement. For the bearing and transducer configuration used in this paper this limit was estimated at 20 μs.

Figure 9 shows limits of operation for the current equipment and bearing geometry. Three regimes have been defined in this graph. When the bearing is operating at a low radial load, then the lubricated contact patch generated can be smaller than the focal spot size of the transducer. Any conditions when the focal spot is larger than the contact size are defined as the "poor operating regime." Note that the lubricated contact size is a function of shaft speed, but as this is not a strong function the line in Fig. 9 appears vertical. For conditions of high shaft speed, a small, fast moving, lubricated contact is generated. If this is such that only one point can be measured per ball then this is also indicated in Fig. 9 as the poor operating regime. If the speed and load are such that the lubricated "contact" region is larger than the spot size of the transducer and two measurement points can be obtained for each ball passage, this was classified as the "acceptable operating regime." In the "good operating regime," more than two reflected signals can be used to characterize each lubricated contact. In this paper all measurements presented have been from within the good operating regime. It is worth noting that an increase in the pulse repetition frequency, a reduction in the focal spot size of the transducer, or an increased test bearing size, will result in an increase in the sizes of the good and acceptable measurement regimes.

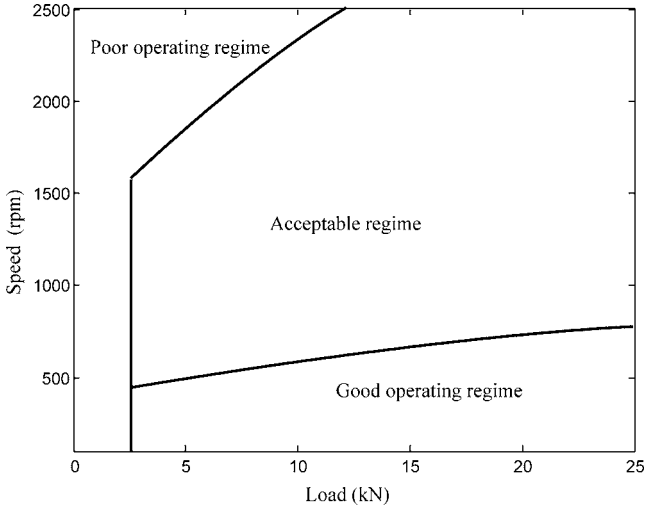


FIG. 9. Limits of operation for lubricant-thickness measurement.

VI. CONCLUSIONS

Ultrasonic reflection coefficient measurements have been used to measure the lubricant-film thickness of a rotating element ball bearing (type 6016). Measurements of lubricant film thickness in the range 0.3–1.0 μm have been made and shown to be in good agreement with known elastohydrodynamic lubrication theory when the radial load was larger than 2.5 kN and the shaft speed was lower than 200 rpm. The ultrasonic transducer produces an "averaging" effect, due to the focal spot size of the transducer, and this was shown to explain the form of the measured lubricant film thickness distribution. In this way the focal spot size defines the spatial resolution of the measurement system. The measurement system was also shown to be limited by the maximum achievable pulse-repetition frequency. This determines the number of measurement points obtained across the lubricated contact. The performance of the experimental system demonstrates that this approach has the potential for condition monitoring of lubricant layers in an industrial application.

ACKNOWLEDGMENTS

This work has been funded by the UK Engineering and Physical Sciences Research Council.

[1]G. W. Stachowiak and A. W. Batchelor, *Engineering Tribology* (Butterworth Heinemann, Woburn, MA, 2001).
[2]B. J. Hamrock, S. R. Schmid, and B. O. Jacobson, *Fundamentals of Fluid Film Lubrication*, Second Edition (Marcel Dekker, New York, 2004).
[3]S. I. El-Sisi and G. S. A. Shawki, "Measurement of oil-film thickness between disks by electrical conductivity," ASME J. Basic Eng. **82**, 12–18 (1960).
[4]K. G. Astridge and M. D. Longfield, "Capacitance measurement and oil film thickness in a large radius disc and ring machine," *Proc. Insti. Mech. Eng.* **182**, 89–96 (1967).
[5]A. Dyson, "Investigation of the discharge-voltage method of measuring the thickness of oil films formed in a disc machine under conditions of elastohydrodynamic lubrication," *Proc. Inst. Mech. Eng.* **181**, 633–645 (1967).
[6]D. O. Ducu, R. J. Donahue, and J. B. Ghandhi, "Design of capacitance probes for oil film thickness measurements between the piston ring and liner in internal combustion engines," J. Eng. Gas Turbines Power **123**, 633–643 (2001).
[7]A. Cameron and R. Gohar, "Theoretical and experimental studies of the oil film in lubricated point contact," Proc. R. Soc. London, Ser. A **291**, 520–536 (1966).
[8]D. A. Richardson and G. L. Borman, "Using fibre optics and laser fluorescence for measuring thin oil films with applications to engines," Society of Automotive Engineers, SAE Paper 912388, 1991.
[9]D. Jiao and J. L. Rose, "An ultrasonic interface layer model for bond evaluation," J. Adhes. Sci. Technol. **5**, 631–646 (1991).
[10]R. E. Challis, T. Alper, A. K. Holmes, and R. P. Cocker, "Near-plane-wave acoustic propagation measurements in thin layers of adhesive polymer," Meas. Sci. Technol. **2**, 59–68 (1991).
[11]V. K. Kinra, P. T. Jaminet, C. Zhu, and V. R. Iyer, "Simultaneous measurement of the acoustical properties of a thin-layered medium: The inverse problem," J. Acoust. Soc. Am. **95**, 3059–3074 (1994).
[12]L. Wang, B. Xie, and S. I. Rokhlin, "Determination of embedded layer properties using adaptive time-frequency domain analysis," J. Acoust. Soc. Am. **111**, 2644–2653 (2002).
[13]W. Anderson, J. Jarzynski, and R. F. Salant, "Condition monitoring of mechanical seals: Detection of film collapse using reflected ultrasonic waves," Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci. **214**, 1187–1194 (2000).
[14]R. S. Dwyer-Joyce, B. W. Drinkwater, and C. J. Donohoe, "The measurement of lubricant-film thickness using ultrasound," Proc. R. Soc. London, Ser. A **459**, 957–976 (2003).
[15]J. Zhang, B. W. Drinkwater, and R. S. Dwyer-Joyce, "Calibration of the ultrasonic lubricant film thickness measurement technique," Meas. Sci. Technol. **16**, 1784–1791 (2005).
[16]S. I. Rokhlin and W. Huang, "Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids: Exact asymptotic-boundary-condition methods," J. Acoust. Soc. Am. **92**, 1729–1742 (1992).
[17]T. Pialucha and P. Cawley, "The detection of thin embedded layers using normal incidence ultrasound," Ultrasonics **32**, 431–440 (1994).
[18]T. A. Harris, *Rolling Bearing Analysis*, 4th ed. (Wiley, New York, 2001).
[19]D. Dowson and G. R. Higginson, *Elasto-Hydrodynamic Lubrication* (Pergamon, New York, 1997).
[20]B. O. Jacobson and P. A. Vinet, "Model for the influence of pressure on the bulk modulus and the influence of temperature on the solidification pressure for liquid lubricants," ASME J. Tribol. **109**, 709–714 (1987).
[21]M. G. Silk, *Ultrasonic Transducers for Nondestructive Testing* (Adam Hilger, Bristol, UK, 1984).
[22]D. P. Orofino and P. C. Pedersen "Angle-dependent spectral distortion for an infinite planar fluid-fluid interface," J. Acoust. Soc. Am. **92**, 2883–2899 (1992).

A phase reconstruction algorithm for Lamb wave based structural health monitoring of anisotropic multilayered composite plates

Jagannathan Rajagopalan,[a] Krishnan Balasubramaniam, and C. V. Krishnamurthy
Center for Nondestructive Evaluation and Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai-600036, India

(Received 4 May 2005; accepted 11 November 2005)

Platelike structures, made of composites, are being increasingly used for fabricating aircraft wings and other aircraft substructures. Continuous monitoring of the health of these structures would aid the reliable operation of aircrafts. This paper considers the use of a Lamb wave based structural health monitoring (SHM) system to identify and locate defects in large multilayered composite plates. The SHM system comprises of a single transmitter and multiple receivers, coupled to one side of the plate that send and receive Lamb waves. The proposed algorithm processes the data collected from the receivers and generates a reconstructed image of the material state of the composite plate. The algorithm is based on phased addition in the frequency domain to compensate for the dispersion of Lamb waves. In addition, small deviations from circularity of the slowness curves of Lamb wave modes, due to anisotropy, are corrected for by assuming that the phase and group velocity directions coincide locally. Experiments were performed on an anisotropic multilayered composite plate containing a single defect. Reconstruction of the defect is carried out using data for a weakly anisotropic Lamb wave mode as a proof of concept of the proposed algorithm. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149775]

PACS number(s): 43.35.Zc, 43.35.Cg, 43.60.Fg [YHB] Pages: 872–878

I. INTRODUCTION

Multilayered composites, in the form of platelike structures, are being increasingly used in variety of structures because of the unique properties they offer. In particular, they are often used for fabricating wings and other substructures of aircrafts. Inspection of these large structures using traditional nondestructive evaluation methods is both time consuming and difficult, especially if the monitoring has to be done *in situ*. Lamb wave based inspection methods have been shown to be well suited for monitoring such structures, since Lamb waves have the capability of traveling large distances without much attenuation. But, Lamb waves are dispersive in nature and hence the velocity of propagation becomes a function of the wave frequency, the thickness of the plate, and the direction of travel (in case the medium is anisotropic). Hence, analyzing Lamb wave data to evaluate the integrity of structures becomes complicated.

Moreover, most of the classical theoretical work on Lamb waves is based on assuming the wave vectors to be one-dimensional.[1] But, in practice the wave vectors need to be two-dimensional. This is because for effective structural health monitoring, it is desirable that the wave interrogates the material at different angles. Also, reflections from a structural feature in the plate may propagate along any direction in the plane of the plate. Hence, it is necessary to incorporate directionality in any inspection procedure in order to image damage in composite materials. The suitability of an array based approach for achieving directionality has been demonstrated in Refs. 2 and 3. The applicability of an array based system for rapid inspection of isotropic plate structures was demonstrated in Refs. 4 and 5, wherein beam steering algorithms were used to process data collected from omnidirectional guided wave transducer arrays. Structural health monitoring of isotropic and anisotropic platelike structures by processing data from multiple transmitters and multiple receivers (MTMR) using tomographic reconstruction techniques has been reported elsewhere.[6–8]

In this paper, we seek to extend the flaw detection technique demonstrated for isotropic plate-like structures in Refs. 4 and 5 to anisotropic materials. In anisotropic plates, such as the fiber reinforced composite laminates, the energy propagation direction is along the preferred "group" velocity direction and can be determined through the analysis of the slowness surface (inverse of velocity profile).[9–14] Anisotropy in solid materials has a tendency to change the direction of the energy of the acoustic wave propagation along preferred orientations. The energy propagation direction at any "phase" velocity angle (wave vector direction) is oriented along the normal to the slowness surface measured at that "phase" angle. Hence, any algorithm for Lamb wave reconstruction in anisotropic materials, unlike in isotropic materials, must consider this dependency of the phase velocity and energy on the direction of wave travel.

The first part of this paper deals with the description of the array system and the method of data acquisition. The second part describes the algorithm for material state reconstruction in anisotropic materials. The final part shows the results of applying the algorithm to experimental data. For

[a]Presently at the Department of Mechanical and Industrial Engineering, University of Illinois Urbana-Champaign.

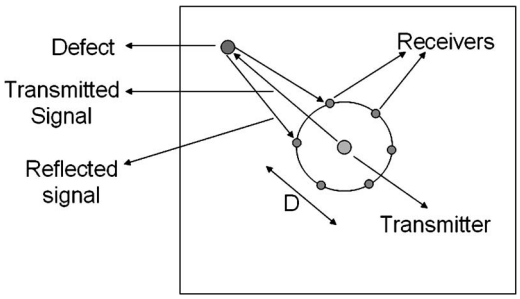


FIG. 1. Architecture for the Lamb wave based SHM system.

applying the algorithm, the dispersion characteristics of the anisotropic material for wave propagation along all directions need to be known *a priori*.

II. ARRAY DESCRIPTION AND DATA ACQUISITION

As indicated in Fig. 1, the array consists of a single transmitter and N number of receivers coupled to one surface of the plate. The receivers are arranged in a circle of diameter D while the transmitter is placed at the center. A polar coordinate system is defined with its origin at the center of the array coinciding with the transmitter position. The positions of the receivers are defined using their radial and angular coordinates. It is assumed that the transmitter and the receivers act as point sources and have equal transmission and reception sensitivity in all directions.

The data acquisition and transformation process begins with the acquisition of the raw data set in the time, **t**, domain. The transmitter is excited and a discretely sampled **t** domain signal is collected at each of the receivers. This data is used to form a matrix, **T**, in which each column represents the raw **t** domain signal collected at each of the N receivers. Consequently, the number of columns in the **T** matrix is equal to the number of receivers. A column wise fast Fourier transform (FFT) is, then, performed on the **T** matrix to transform it to a matrix, **F**, containing the complex spectra in the frequency, $\boldsymbol{\omega}$, domain. The transformation to the frequency domain is done so that the effects of guided wave dispersion can be taken into account.

III. PHASE RECONSTRUCTION ALGORITHM

The purpose of the phase reconstruction algorithm is to determine the locations from which the waves sent by the transmitter are reflected back. In the algorithm, the reflected wave packets are assumed to come from a particular direction (angle) and the signal received at each receiver is shifted appropriately (in the $\boldsymbol{\omega}$ domain) and added. If the assumed direction is the actual direction of the reflected wave then this process results in a coherent addition. This procedure is repeated for a set of equi-spaced angles that span the entire 360°. The phase-reconstructed signals are then transformed into the wave number domain, **k**, using interpolation. The locations from which the waves are reflected back are then obtained by performing an IFFT on these phase-reconstructed signals.

As mentioned earlier, algorithms for beam steering from omni-directional transducer arrays, using phased addition in the **k** domain have been proposed in Refs. 4 and 5. But such algorithms are restricted to isotropic materials where a transformation from **t** domain to $\boldsymbol{\omega}$ domain and subsequently to **k** domain is possible. In anisotropic materials, the wave number is a function of both the wave frequency and the direction of wave propagation. Hence, a transformation from the **t** domain to **k** domain is not possible when the direction of wave propagation is not known, as is the case with reflections from unknown defects.

In the first step of the algorithm, a particular guided wave mode is chosen and phased addition of the columns of the **F** matrix is performed along n equispaced reconstruction angles using the dispersion characteristics of the mode. These reconstruction angles, represented by θ_p, are the assumed angles of propagation of the reflected waves. This phased addition in the $\boldsymbol{\omega}$ domain leads to a matrix, **Q**, in which the rows represent different frequencies, ω_k, and the columns represent different reconstruction angles, θ_p. The elements in **Q** are calculated using the following expression:

$$\mathrm{Q}_{\omega p} = \sum_{j=1}^{N} f_{\omega j} \exp(-i2\pi \mathrm{k}(\omega_k, \theta_p)\mathrm{x}_{pj}), \tag{1}$$

where

$$\mathrm{x}_{pj} = \mathrm{R}_j \cos(\Phi_j - \theta_p). \tag{2}$$

In the above expressions R_j and Φ_j are the polar coordinates of the jth receiver, while x_{pj} represents the change in path length required for the signal received at the jth receiver to maintain coherence along θ_p.

In making the phase corrections to $\mathrm{f}_{\omega j}$'s in Eq. (1), the algorithm assumes that the phase velocity and group velocity directions are the same locally. For the spatially tight transmit-receive configuration considered, and for a flaw in the far-field where the algorithm is intended to be applied, the ray paths from the transmitter to the flaw and from the flaw to the receivers can be taken to be nearly equal along the given orientation based on the principle of reciprocity. Accordingly, the algorithm is expected to be applicable for weakly anisotropic materials over all angles, for moderate anisotropic materials over angular sectors around symmetry axes and for strongly anisotropic materials along the symmetry axes. For the algorithm to be applicable even for strongly anisotropic materials, one has to first obtain the phase velocity angle, Ψ_p, corresponding to the group velocity angle θ_p and then use $\mathrm{k}(\omega_k, \Psi_p)$ for making the phase correction.

Proceeding to the next step of the algorithm, the dispersion characteristics of the chosen guided wave mode are used for the different directions, θ_p, and each column in **Q** is separately interpolated to obtain a matrix, **W**, where each column contains spectra at points equispaced in the **k** domain. The interpolation procedure for single column is performed as follows. Let $\omega_1, \omega_2, \ldots, \omega_m$ be the m discrete equispaced frequencies corresponding to the m rows in **Q**, with ω_1 as the minimum frequency and ω_m as the maximum. For a particular column, representing a particular angle, let $\mathrm{k}_1, \mathrm{k}_2, \ldots, \mathrm{k}_m$ be the wave numbers corresponding to the guided wave mode of interest ($\mathrm{k}_1, \mathrm{k}_2, \ldots, \mathrm{k}_m$ are known since both the frequency and the angle of wave propagation are known).

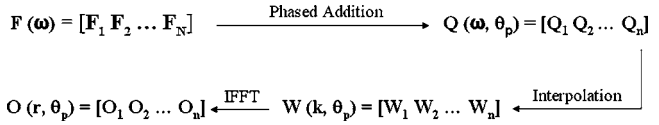


FIG. 2. Phase reconstruction and inversion procedure for Lamb waves in anisotropic materials.

Hence, the spectra at $k_1, k_2, \ldots, k_m$ is known. But, the wave numbers $k_1, k_2, \ldots, k_m$ are not equispaced since the phase velocity changes with frequency. Hence, m equispaced wave numbers, $k_1', k_2', \ldots, k_m'$, are chosen between k_1 and k_m with $k_1'(=k_1)$ as the minimum and $k_m'(=k_m)$ as the maximum. Since the spectra at $k_1, k_2, \ldots, k_m$ is known, the spectra at k_1', $k_2', \ldots, k_m'$ can be obtained by interpolation.

To obtain the output matrix, **O**, in the polar domain, the columns in **W** are subject to an inverse fast Fourier transform (IFFT) to convert them from the $\mathbf{k}$-$\boldsymbol{\theta}$ domain to the $\mathbf{r}$-$\boldsymbol{\theta}$ domain. Now each element in **O** has a corresponding radial and angular position attached to it. Reconstruction is obtained by plotting the amplitudes of the elements in **O** as a function of their polar position. The procedure for phase reconstruction is summed up in Fig. 2.

The effectiveness of this algorithm for imaging the flaw in the "near field" is limited by the assumption made during the phase reconstruction. Here, the "near field" is defined to be approximately 2–3 times the diameter of the sensor configuration (D). The phase correction algorithm effectively approximates the path lengths between the flaw and the receiver to be equal to the projection along the angle that is being reconstructed.[4,5] While this assumption is valid in the "far field," this approximation does lead to distortions in the flaw imaging in the "near field" region.

Also, it must be noted that as the sensor configuration diameter D increases, the phase resolution in the "far field" increases leading to improved reconstruction. However, the increase in D will result in an increase in the "near field" region. Hence, the diameter D is a critical parameter that must be carefully chosen depending on the type of structure and the region of critical flaw location.

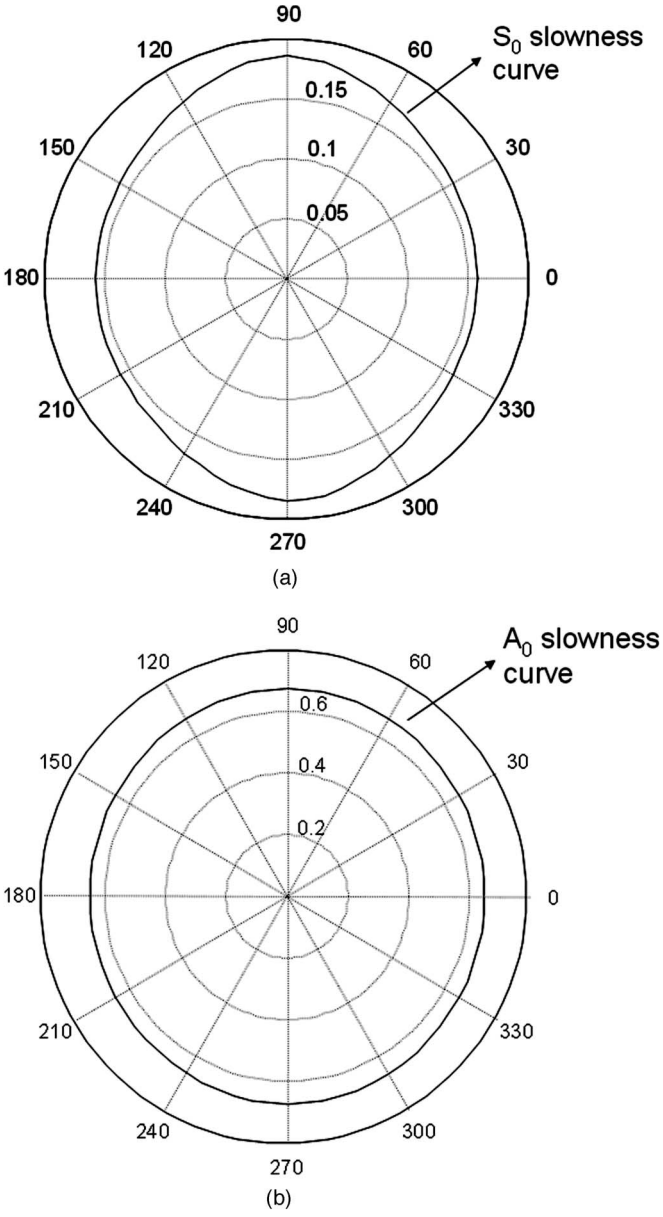


FIG. 3. Slowness curves of (a) S_0 mode and (b) A_0 mode for 3.15 mm thick multilayered composite plate.

IV. EXPERIMENTS AND RESULTS

Two experiments, the first to reconstruct the edges of a composite plate, and the second to reconstruct a defect (hole) in the composite plate, were performed to validate the algorithm. The plate, 3.15 mm thick, was made of 21 layers of unidirectional composite, each 0.15 mm thick. The fiber orientation in the first 7 layers was as follows: +45°, −45°, 0°, 90°, 0°, −45°, and +45°. The same pattern was repeated in the next 14 layers. To obtain the dispersion curves, the averaged properties of the material were used. Using the engineering constants of the unidirectional composite, the elastic coefficients were found and the elastic coefficients matrix was formed. Then the effective elastic coefficients matrix for different orientations (+45°, −45°, 90°) was found by transforming the elastic coefficient matrix in the respective direction. The elastic coefficients matrix for each orientation was then multiplied by the number of layers (since each layer is of same thickness) in that particular orientation and added together. This matrix was divided by the total number of layers to get the "effective" elastic coefficients matrix.[15,16] The "effective" elastic coefficients matrix was used to obtain the dispersion curves in different directions using the DISPERSE™ software (developed by Imperial College, UK). Typical slowness curves for S_0 mode, used in the reconstructions, and A_0 mode are shown in Figs. 3(a) and 3(b), respectively. As can be seen from the figures, the slowness profile of the S_0 mode shows a fair amount of variation with angle while that of the A_0 mode is relatively uniform. The radii of the circles in Fig. 3 are in ms/m.

In both the experiments, a Panametrics 5058PR broadband Pulser/Receiver was used to excite a Panametrics Videoscan 500 kHz Transducer, which was used as a transmitter. A similar transducer was used as receiver. An Agilent 54621A oscilloscope was used for signal acquisition. The receiver was kept at 18 equispaced locations on a circle, of

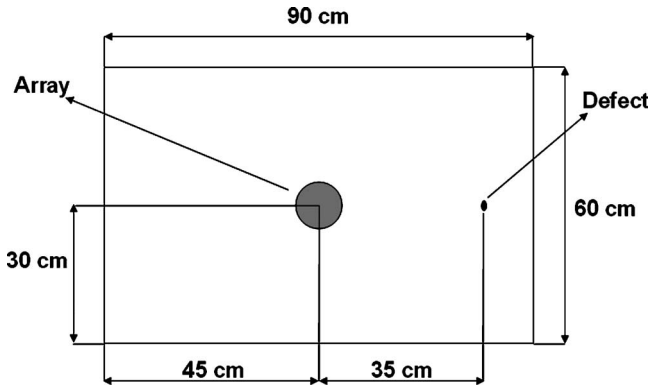


FIG. 4. Plate dimensions, array, and defect locations for the experiments.

10 cm diameter with the transmitter at the center, and signals were collected at each of these locations to simulate an array with 18 receivers. The transmitter was excited with a 100 V signal with center frequency of 500 kHz. The signals were sampled at 2.5 MHz and filtered after acquisition using a band pass filter of 100–1000 kHz range to eliminate noise and undesirable low frequency components. The dimensions of the plate and locations of the array and the defect (for the second experiment) are shown in Fig. 4.

Figure 5 shows the reconstruction after applying the algorithm, with 36 reconstruction angles. The gray scale on the plot is logarithmic with a range of 12 decibel (dB). The array is marked at the center and the edges can be seen at the appropriate locations as black patches, the black patches indicating higher amplitudes obtained as a result of the coherent addition of the reflected signals obtained at the various receivers. When the dB range of the reconstruction is increased, a few artifacts show up. Figure 6 shows the reconstruction with the same experimental data, assuming the plate to be isotropic, with the dispersion characteristics for $\theta=0°$ propagation being used as the dispersion characteristics for all directions. In this reconstruction, only the edges along $\theta=0°$ and $\theta=180°$ are located accurately, indicating the need to incorporate the anisotropic correction during reconstruction.

A typical waveform obtained at a receiver, comprising both the transmitted and reflected signals, is shown in Fig. 7.

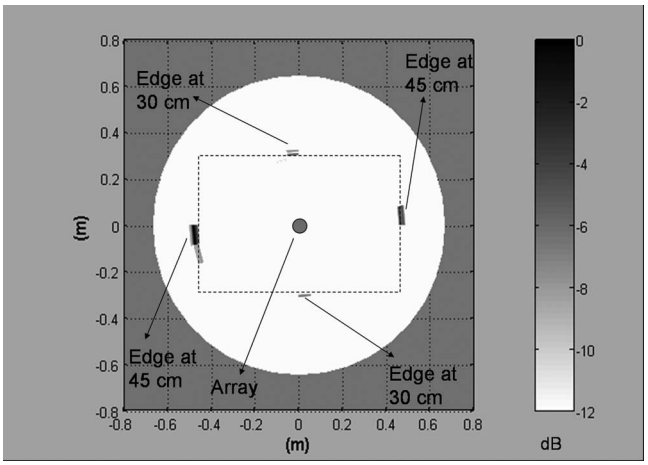


FIG. 5. Reconstruction of edges of composite plate with 12 dB range.

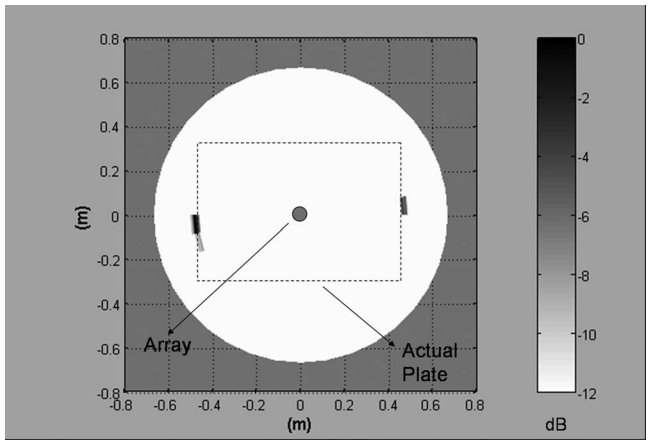


FIG. 6. Reconstruction of edges, assuming the plate to be isotropic, with the dispersion characteristics for $\theta=0°$ being used as the dispersion characteristics for all directions.

Figures 8(a) and 8(b) show the frequency content of the transmitted and reflected signals. While the transmitted signal has a large frequency content around the excitation frequency at 500 kHz, the reflected signals have most of their frequency content around 200 kHz. Since only the S_0 and A_0 modes exist at these frequencies and S_0 is the faster among them and did not show much attenuation, it was chosen for reconstruction. Another reason for using S_0 was that it seemed to undergo no mode conversion at the edges. We suspect that the frequencies around 500 kHz in the transmitted signal get attenuated because the dominant S_0 mode is highly dispersive around that frequency.

In the second experiment a roughly oval-shaped hole, 0.4 cm wide and 0.8 cm long, was drilled in the 0° direction at a distance of 35 cm from the center of the plate (Fig. 4), to induce a defect in the plate. The flaw was chosen to simulate a through hole that can occur due to high energy impacts on aircraft wings, or even moderate impacts, if they occurred at locations that already had some internal defect, like a void. As the technique is intended for locating damage in large plates, the edge-effects arising out of using a small plate had to be kept to a minimum. The flaw location was chosen such that the interference to the signal reflected from the flaw

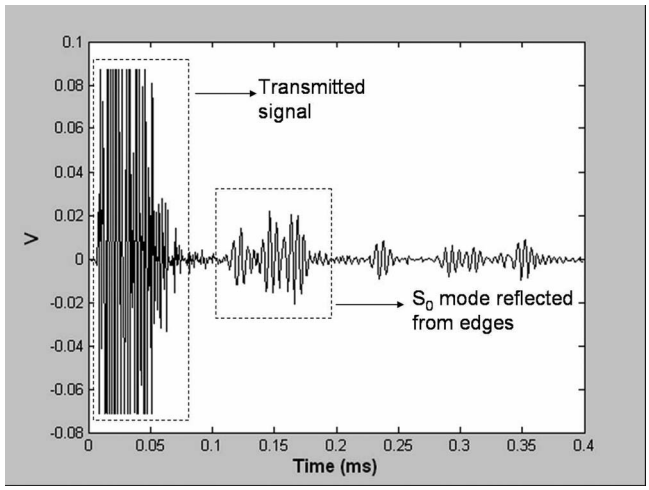
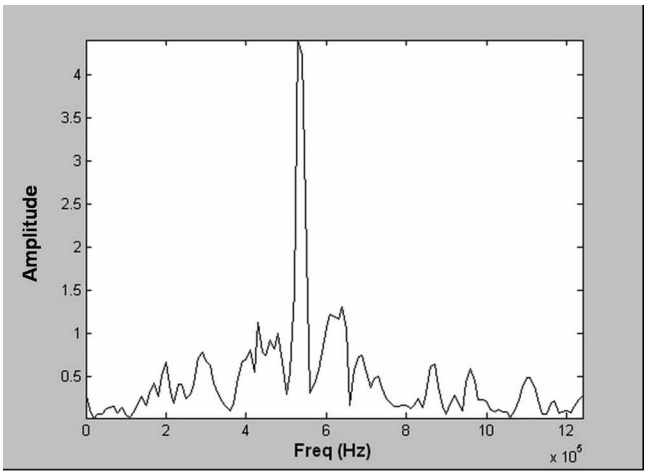
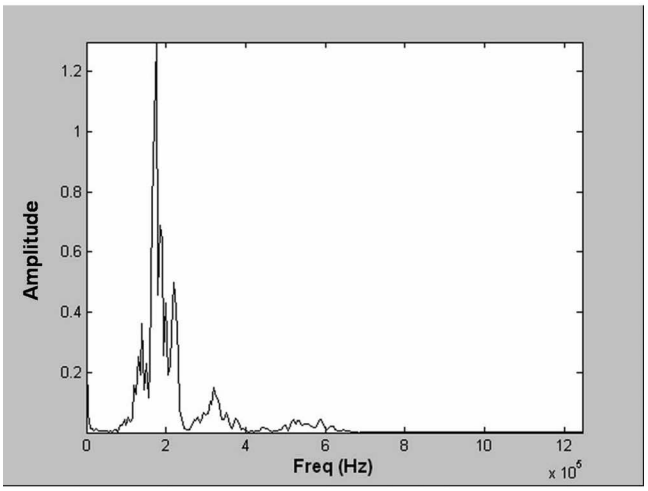


FIG. 7. A typical waveform obtained at a receiver.



(a)



(b)

FIG. 8. Frequency contents of (a) transmitted signal and (b) reflected signal.

from edge and corner reflections as well as the transmitted signal was minimal. The distance of the flaw from the array was chosen such that it was sufficiently far off to avoid any "near field" effects. The data was collected in the same manner as in the first experiment. Figure 9 shows the reconstruction with a range of 20 dB, the reconstruction being presented with a higher dB range so that the defect can be seen more clearly.

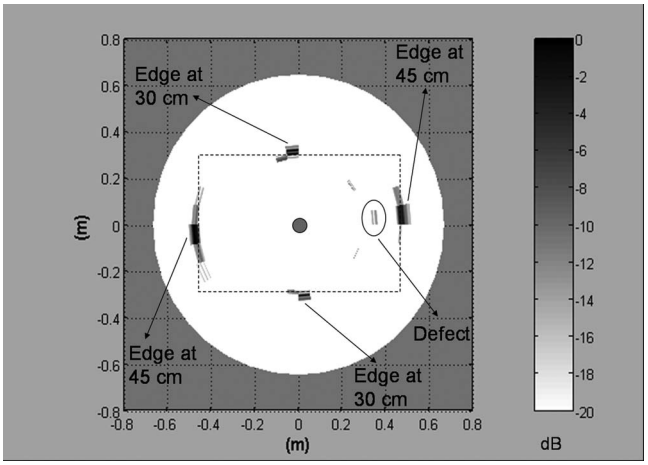


FIG. 9. Reconstruction of defect (hole) in the composite plate with 20 dB range.

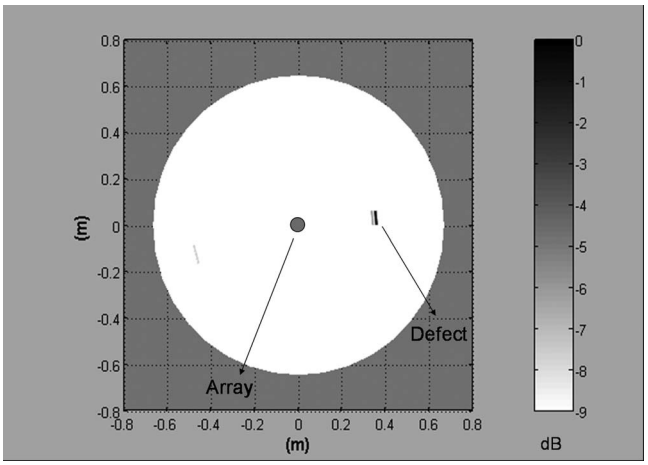


FIG. 10. Reconstruction of defect using difference of amplitudes, with 9 dB range.

While the presence of the defect can be made out in Fig. 9, the clarity is not very good since the reflection from the edges dominate over the reflection from the defect. Therefore, to get a better reconstruction, the output matrix **O** (containing the amplitude at different locations) obtained from the first experiment (when the plate was defect less) was subtracted from the output matrix **O** of the second experiment and the difference was plotted in the logarithmic scale. This was done so that the reflection from the edges cancelled out leaving only the reflection from the defect. Figures 10 and 11 show the logarithmic plot of the difference of the **O** matrices, $\mathbf{O}'$, with 9 dB and 12 dB range, respectively. A comparison of Figs. 9 and 10 reveals that the use of $\mathbf{O}'$ as a reconstruction parameter does indeed leads to the suppression of the edge reflections to a very good extent, though it must be noted that Fig. 10 has a lower dB scale. But, even if the plot is on a 12 dB scale (which has been used throughout), as in Fig. 11, the edge reflections are much weaker as a comparison with Fig. 5 would readily reveal. If the reconstruction is plotted with even higher dB ranges, a few other artifacts, mainly due to noise in the signals, do show up. Hence, a calibration procedure may be necessary in order to reduce the number of false alarms during SHM. All the reconstructions for the second experiment were obtained with 36 reconstruction angles.

The analytical method for obtaining the dispersion curves can be used if the material properties and the ply lay-up for the composite are available. Alternatively, if such information is not available, the single-transmitter-multiple-receivers (STMR) array allows for a self-calibration mode for obtaining the dispersion function, since the receivers first receive the transmitted wave. In the self-calibration mode, the anisotropy and the dispersion properties, as function of orientation, can be experimentally derived from the first received signal at each of the receivers. The wave velocities of the individual modes can be obtained by measuring the time difference between the trigger and the arrival time of the received signal at each of the receiver. If sufficient numbers

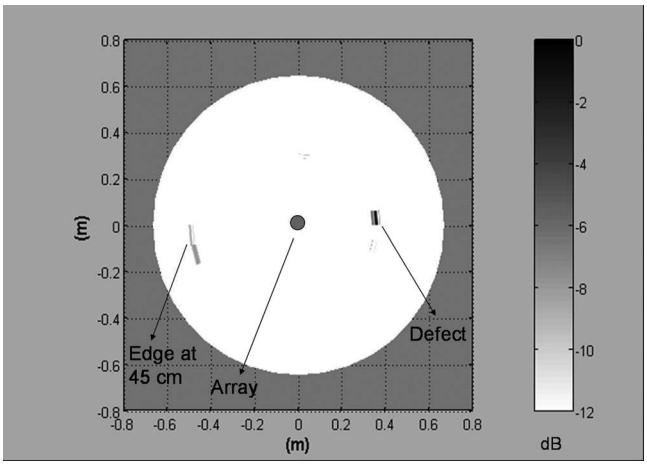


FIG. 11. Reconstruction of defect using difference of amplitudes, with 12 dB range.

of velocity are available, the elastic moduli that define the wave propagation can be obtained by using inverse techniques that have been documented.[17–20] Once the elastic moduli are obtained, the dispersion curve for the fundamental modes can be computed. Another technique for obtaining the dispersion behavior of the as received Lamb wave mode would be to utilize the time-frequency transforms such as wavelets.[21,22] Once the dispersion characteristics, as a function of orientation, have been experimentally obtained using the self-referencing mode, they can be then used in the phase reconstruction algorithm in the SHM mode. However, for this to be reliable the number of receivers must be sufficiently high.

V. SUMMARY

In this paper, a method for structural health monitoring (SHM) of anisotropic platelike structures using a STMR array of transducers has been developed. The method uses a new phase reconstruction algorithm, which takes into account the directional dependence of the Lamb wave dispersion characteristics in anisotropic plates, for processing data obtained from the STMR array. A system which is conceptually similar to the STMR array, but using only a single PZT transmitter and receiver, has been shown to be capable of locating damage of dimensions much smaller than the plate dimension with good accuracy. The phase reconstruction algorithm works well in the "far field," but may show distortions/artifacts in the "near field" which is the region within approximately 2–3 times the sensor configuration diameter (D). Also, more experiments, with different flaws, need to be performed to assess the efficacy of this technique to image other types of defects.

It has been demonstrated elsewhere[5] that a multiple-transmitter-multiple-receiver array system provides a better signal to noise ratio, which results in images with improved resolution and clarity for plates made from isotropic materials. But a STMR system was chosen here since it requires simpler electronics and a reduced footprint (desired during the SHM of aerospace structures). Also, other types of piezomaterials like PVDF film sensors (which have good reception capabilities but poor transduction capabilities) can be used as receivers in an STMR system along with a PZT transmitter. Other options include MEMS and Fiber Optics based sensor systems. This will result in a much lighter and a more compact array system.

Several other array based inspection techniques have been used by researchers to increase the signal to noise ratio so that better quality images of the damage location can be obtained. Two such techniques are based on beam forming using "phased arrays" and time-reversal concepts, demonstrated for example in Refs. 23 and 24, respectively. However, for the time-reversal concept, several issues, like accounting for group velocity change with direction, still needs to be resolved. Another technique used to identify damage in platelike structures has been by measuring changes in its dynamic response to external excitations.[25] An ultrasonic pulse is sent through the structure and an array of sensors is used to measure some physical parameter like displacement, velocity, etc., at selected points. The physical quantity measured by the sensors changes if the structure suffers any damage and this change can be used to identify and locate the damage. While this technique offers the advantage of a fairly sparse sensor configuration, it does not provide accurate localization of the damage. The success of this technique also depends to some extent on selecting the right parameter.

The technique demonstrated in this paper to image features in the structure, such as voids and/or delaminations, that reflect guided waves back to several or all the receivers, offers good localization of the defect, has a compact and small footprint for the sensor configuration, is easy to implement, and is applicable to anisotropic materials. For anisotropic materials which are not well characterized, i.e., the elastic properties and ply orientations are not available *a priori*; the STMR array can be designed to be self-calibrating by using the initial signal received by the receivers to measure the velocity and energy distribution profiles. These profiles lead to the evaluation of the dispersion relationships that can subsequently be employed for the phase reconstruction algorithm. STMR configurations, other than the circular configuration considered here, can be tailored to account for the anisotropy of the material, to improve the SHM of composite structures.

ACKNOWLEDGMENTS

This work was funded by the Aeronautical Development Agency, Bangalore through the DISMAS program.

[1] I. Victorov, *Rayleigh and Lamb Waves: Physical Theory and Applications* (Plenum, New York, 1967).
[2] R. Sicard, J. Goyette, and D. Zellouf, "A SAFT algorithm for Lamb wave imaging of isotropic platelike structures," Ultrasonics **39**, 487–494 (2002).
[3] P. Wilcox, M. Lowe, and P. Cawley, "Lamb and SH wave transducer arrays for the inspection of large areas of thick plates," in *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, New York, 1999), Vol. 19A, pp. 1049–1056.

[4]P. D. Wilcox, "Guided wave beam steering from omnidirectional transducer arrays," in *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, New York, 2002), Vol. 22A, pp. 761–768.
[5]P. D. Wilcox, "Omnidirectional guided wave transducer arrays for the rapid inspection of large areas of plate structures," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**(6), 699–709 (2003).
[6]S. Mahadev Prasad, R. Jagannathan, Krishnan Balasubramaniam, and C. V. Krishnamurthy, "Structural health monitoring of anisotropic layered composite plates using guided ultrasonic Lamb wave data," in *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, New York, 2003), Vol. 23B, pp. 1460–1467.
[7]S. Mahadev Prasad, Krishnan Balasubramaniam, and C. V. Krishnamurthy, "Structural health monitoring of composite structures using Lamb wave tomography," Smart Mater. Struct. **13**, N73-N79 (2004).
[8]R. Jagannathan, B. V. Somasekhar, Krishnan Balasubramaniam, and C. V. Krishnamurthy, "Plate waves structural health monitoring of composite structures," *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, New York, 2005), Vol. 24B, pp. 1802–1808
[9]A. H. Nayfeh, *Wave Propagation in Layered Anisotropic Media* (North-Holland, Amsterdam, 1995).
[10]B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Malabar, 1990), Vols. 1 and 2.
[11]R. Sullivan, K. Balasubramaniam, and A. G. Bennett, "Plate wave flow patterns for ply orientation imaging in fiber reinforced composites," Mater. Eval. **54**(4), 518–523 (1996).
[12]K. Balasubramaniam and Y. Ji, "Guided wave analysis in inhomogeneous plates," in *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, New York, 1995), Vol. 14, pp. 227–234.
[13]K. Balasubramaniam and Y. Ji, "Influence of skewing on the acoustic wave energy vector behavior in anisotropic material systems," J. Synchrotron Radiat. **236**(1), 166–175 (2000).
[14]S. Baly, C. Potel, J.-P. de Belleval, and M. Lowe, "Numerical and experimental deviation of monochromatic Lamb wave beam for anisotropic multilayered media," *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, AIP Proceedings, CP615, Vol. 21, 270–278 (2002).
[15]T. D. Lhermitte and B. Perrin, "Anisotropy of the elastic properties of cross-ply fiber-reinforced composite materials," Proc.-IEEE Ultrason. Symp., pp. 825–830 (1991).
[16]John J. Ditri and Joseph L. Rose, "An experimental study on the use of static effective modulus theories in dynamic problems," J. Compos. Mater. **27**(9), 934–943 (1993).
[17]R. A. Kline, *Nondestructive Characterization of Composite Media* (Technomic, Lancaster, 1992).
[18]J. J. Ditri, "On the determination of the elastic moduli of anisotropic media from limited acoustical data," J. Acoust. Soc. Am. **95**(4), 1761–1767 (1994).
[19]K. Balasubramaniam and N. S. Rao, "Inversion of composite material elastic constants from ultrasonic bulk wave phase velocity data using genetic algorithms," Composites, Part B **29**(B), 171–180 (1998).
[20]K. Balasubramaniam, "Inversion of ply lay-up sequence for multi-layered fiber reinforced composite plates using a genetic algorithm," Nondestr. Test. Eval. **15**, 311–331 (1999).
[21]K. L. Veroy, S. C. Wooh, and Y. Shi, "Analysis of dispersive waves using wavelet transform," in *Review of Progress in Quantitative Nondestructive Evaluation* (American Institute of Physics, Melville, New York, 1999), Vol. 18, pp. 687–694.
[22]M. Niethammer, L. J. Jacobs, J. Qu, and J. Jayrzynski, "Time frequency representation of Lamb waves using reassigned spectrogram," J. Acoust. Soc. Am. **107**, L19–L24 (2000).
[23]Shankar Sundararaman, Douglas E. Adams, and Elias J. Rigas, "Structural damage identification in homogeneous and heterogeneous structures using beamforming," Structural Health Monitoring: An International Journal **4**(2), 171–190 (2005).
[24]Chun H. Wang, James T. Rose, and Fu-Kuo Chang, "A synthetic time-reversal imaging method for structural health monitoring," Smart Mater. Struct. **13**, 415–423 (2004).
[25]Ajit Mal, Fabrizio Ricci, Sauvik Banerjee, and Frank Shih, "A conceptual structural health monitoring system based on vibration and wave propagation," Structural Health Monitoring: An International Journal **4**(3), 283–293 (2005).

A comparison of the underwater acoustic performance of single crystal versus piezoelectric ceramic-based "cymbal" projectors

James F. Tressler
Naval Research Laboratory, Washington, D.C. 20375

Thomas R. Howarth and Dehua Huang
Naval Sea Systems Command, Division Newport, Newport, Rhode Island 02841

(Received 30 June 2005; revised 16 November 2005; accepted 16 November 2005)

For nearly 50 years, piezoelectric ceramics (primarily from the PZT family) have been the materials of choice as the active elements in underwater electroacoustic sound projectors. There is currently great interest in the materials science community for the consideration of newly discovered single crystal relaxor ferroelectric compositions as a potential replacement in applications that utilize piezoelectric ceramics. One of the salient features of single crystal ferroelectrics is piezoelectric coefficients that are three to seven times greater than those found in PZT. Most of the single crystal data reported in the open literature, however, are for near static drive conditions. This paper reports on the acoustic performance of prototype underwater sound projectors built from single crystal materials and driven at high drive levels over the frequency range of 500 Hz to 25 kHz. It is shown that the single crystal-based projectors exhibit at least a 4 dB higher source level as compared to identical PZT-based units. In addition, the volt-amp product required to produce 1 W of acoustic output is approximately one-third as much. It has also been demonstrated that when driving either the PZT-based or single crystal projectors for 1–2 h under high drive, no significant degradation in acoustic performance occurs. [DOI: 10.1121/1.2150153]

PACS number(s): 43.38.Fx, 43.38.Hz, 43.30.Yj [AJZ] Pages: 879–889

I. INTRODUCTION

There has been considerable interest recently in the materials science community in the potential use of single crystal ferroelectrics as a substitute for conventional lead zirconate titanate (PZT) piezoelectric ceramics in a variety of applications.[1–5] It has been shown[1,6] that single crystals of some relaxor-type[7] ferroelectric compositions can generate strains (under a near dc electric field) in excess of 1% and exhibit electromechanical coupling coefficients greater than 90%. These compare with strains on the order of 0.1% and coupling coefficients of ~70% seen in conventional PZT ceramics.[8] Additionally, these single crystal materials possess piezoelectric coefficients three to seven times greater than those found in PZT.[9–11] If this significant improvement in materials properties can be transferred to devices such as underwater sound projectors, a marked enhancement in acoustic performance should be expected.

Single crystal piezoelectrics are not a new phenomenon. They were, in fact, the very first of the electroactive materials to be used in underwater electroacoustic projectors.[12] Quartz, Rochelle salt, tourmaline, ammonium dihydrogen phosphate, and lithium sulfate monohydrate were the major players from the end of World War I until the mid-1940s.[13] The major problems associated with the use of these single crystal materials are low piezoelectric activity, instability, or solubility in water. With the exception of a few specialty applications,[14] the use of these single crystal materials was superseded with the discovery of the ceramic (i.e., polycrystalline) compositions barium titanate ($BaTiO_3$) in the early 1940s, and $Pb(Zr_{0.52}Ti_{0.48})O_3$ (aka PZT) in the mid-1950s. The PZT family of compositions has been the workhorse of the piezoelectric community since the early 1960s.

An additional type of single crystal materials was discovered in the late 1990s[6,15] and is now commercially available. These are single crystals of relaxor-type ferroelectrics having the following compositions:

$$\text{PZN-}y\text{PT} = [\text{Pb}(\text{Zn}_{1/3}\text{Nb}_{2/3})\text{O}_3]_{(1-y)} \times [\text{PbTiO}_3]_y; \quad y \text{ ranges from 4.5 to 8 \%}, \tag{1}$$

$$\text{PMN-}x\text{PT} = [\text{Pb}(\text{Mn}_{1/3}\text{Nb}_{2/3})\text{O}_3]_{(1-x)} \times [\text{PbTiO}_3]_x; \quad x \text{ ranges from 28 to 33 \%}. \tag{2}$$

Although single crystal materials are not polycrystalline like PZT ceramics, they nonetheless still have a multidomain structure. As such, they require poling under a large static electric field, just like PZTs, prior to being used. It is by the appropriate engineering of the single crystal domain states, via crystal orientation and poling direction, that these single crystal materials achieve their superior piezoelectric properties.[3,6]

The piezoelectric performance of these single crystal materials is affected by chemical composition, crystal orientation, and temperature.[6,9] There are some characteristics associated with these single crystals (see Table I) that a transducer designer needs to take into consideration. One is a very

TABLE I. Properties of piezoelectric materials.

Material	Y_{11} ($\times 10^{10}$ N/m^2)	d_{31} ($\times 10^{-12}$ m/V)	K_{33}	E_c (kV/cm) at ~1 Hz	T_{max}
PZT-4	8.2	−135	1300	18.8	300 °C$\approx T_c$
PZT-5H	6.2	−250	2900	10.8	190 °C$\approx T_c$
PMN-30%PT	1.9	−921	5700	2.4	90 °C$\approx T_{R\text{-}T}$
Reference	8 and 11	8 and 11	measured	measured	8 and 16

low electric coercive field E_c, which means that the material normally needs to be electrically biased so it will not depole during operation. Another is a lower use temperature range than PZT piezoelectric ceramics. The maximum operating temperature for a piezoelectric ceramic is taken to be just below its Curie temperature, which is the temperature at which the crystal structure reverts to cubic and the piezoelectricity is lost due to crystallographic symmetry constraints. Although single crystals likewise have a Curie temperature associated with them, they also have an additional crystallographic phase transition that occurs between roughly 80 and 100 °C (depending upon the chemical composition). At this rhombohedral to tetragonal phase change, a marked drop in piezoelectric activity occurs.[3,16] It has also been demonstrated[17] that when driving PMN-PT single crystals in a tonpilz projector at its resonance frequency under an electric field of 120 V_{rms}/mm, in less than 30 min the impedance decreased by 40% and a >45 °C rise in temperature was measured when increasing the duty cycle from 1% to 7.5%.

Most published single crystal property data is for unloaded samples driven at low (near-static) frequency. This paper will present the performance capability of single crystals under high drive conditions from 0.5–25 kHz, where the crystals have been incorporated into "cymbal"-type flextensional drivers in a 7.6 cm × 7.6 cm × 0.76 cm underwater acoustic projector subassembly. This subassembly is in fact one of twelve individual modules that comprise the full projector design configured to operate inside an autonomous underwater vehicle (AUV).[18] This study will conclude by comparing the performance of the single crystal projector to identical projectors made from PZT-4 and PZT-5H.

II. PROJECTOR DESIGN AND FABRICATION

A. Cymbal elements

The cymbal[19] is a miniature Class V flextensional transducer.[20] A cross-sectional sketch, accompanied by a photograph of the cymbal driver used in our projectors is shown in Fig. 1. An individual cymbal element consists of a piezoelectric ceramic or single crystal disk, a pair of studded metal caps, and an adhesive that bonds the caps to the piezoelectric disk. Each piezoelectric disk measures 12.7 mm in diameter and is 1.0 mm thick. Electrodes are on each face of the disk. For the single crystal the electrode is sputtered-on gold. It is baked-on silver for the ceramic. The disk is poled in its thickness direction.

Cymbals operate by transforming the small radial displacement and vibration velocity of the piezoelectric disk into a much larger displacement and vibration velocity normal to the surface of the endcaps. The endcaps serve to mechanically amplify the in-air displacement of the piezoelectric element by an order of magnitude below the fundamental cymbal resonance frequency[21] and by over two orders of magnitude at the cymbal resonance frequency.[22]

The cymbals were made using three different active materials: single crystal, PZT-5H, and PZT-4 piezoelectric ceramic. The single-crystal material selected for this study was PMN-30%PT. It is a multidomain single crystal poled along the [001] crystallographic direction, which is off the polarization direction of ⟨111⟩. Its d_{31} coefficient is over 3.5 times that of PZT-5H piezoelectric ceramic and almost 8 times greater than PZT-4 (see Table I). The relevant electromechanical properties for these three different piezoelectric materials are compared in Table I. It should also be noted that the crystallographic point group (i.e., symmetry elements) of a poled piezoelectric ceramic disk and a single crystal disk are not the same. A poled piezoelectric ceramic disk has

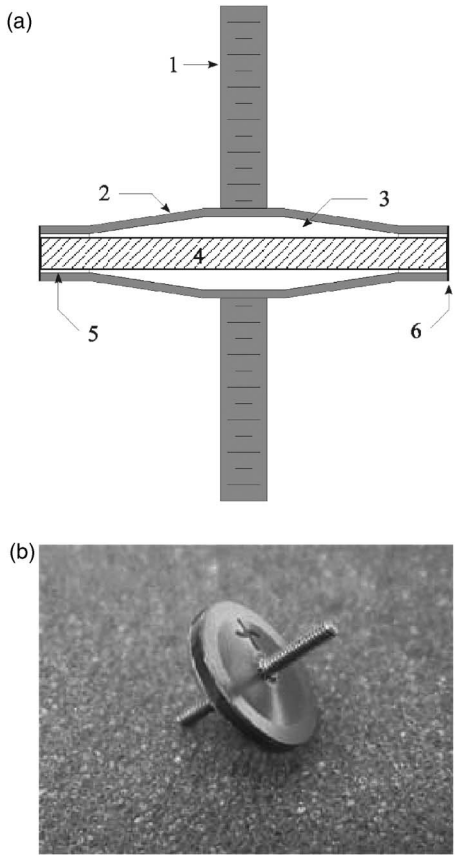


FIG. 1. Two perspectives of the cymbal element (a) cross-sectional sketch and (b) photograph, where the numerals in (a) designate: 1-stud, 2-cymbal cap, 3-air cavity, 4-piezoelectric disk, 5-adhesive layer, and 6-Conathane coating.

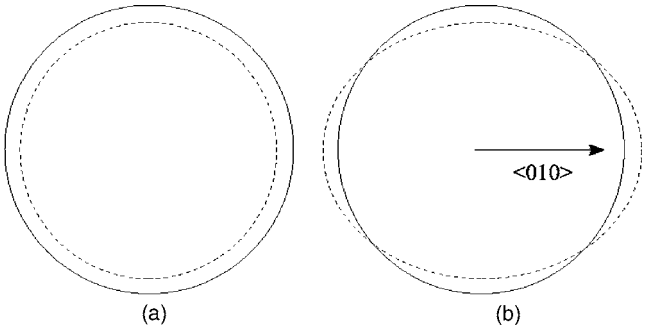


FIG. 2. Sketch showing the relative radial motion of a through-the-thickness poled (a) PZT disk and (b) single crystal disk. The ⟨010⟩ direction in the crystal is indicated by the arrow.

"∞m" symmetry, i.e., its piezoelectric properties are isotropic in the planar direction. Its radial motion is as seen in Fig. 2(a). A poled single crystal disk, on the other hand, exhibits radial motion that is more elliptical in shape [Fig. 2(b)]. This elliptical motion is a consequence of the anisotropy in the single crystal domain structure which drops the symmetry below the expected mm2.

The caps, along with the studs, are made from titanium. Titanium was selected because it gives the best cymbal force/displacement characteristics for this application. The geometry and dimensions of the caps are shown in Fig. 3 and Table II, respectively. The studs, in conjunction with the endcaps, serve as the electrical conduit from the piezoelectric material to the electrical leads. An electrically conductive adhesive is used to bond the caps to the piezoelectric. The adhesive used to bond the PZT piezoelectric ceramic to the caps is a one component film which has uniform geometry and dimensions. Its cure temperature is 120 °C. Because of the ~90 °C rhombohedral to tetragonal phase transition in the single crystal material, a different conductive adhesive (two-part epoxy) was used which has a cure temperature of 80 °C. This adhesive paste had to be applied by hand using an artist's knife. Consequently, it is difficult to maintain a uniform bond width and bond thickness. After the caps were bonded to the ceramic, a thin layer (~150 μm) of Conathane® was applied to the outside rim of the ceramic as protection against electrical breakdown across the air gap between the metal caps.

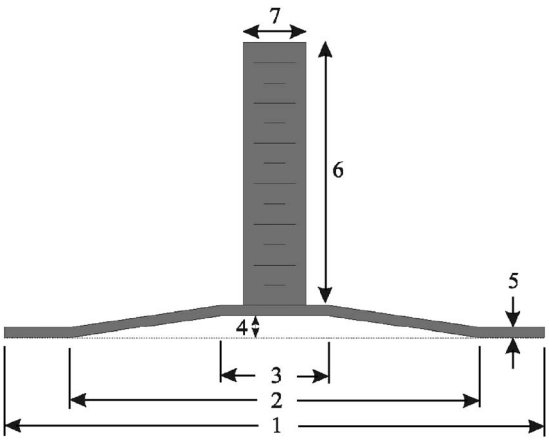


FIG. 3. Cross-sectional view of the cymbal cap. Numerals correspond to parameters in Table II.

TABLE II. Dimensions of the cymbal endcaps.

Parameter	dimension (mm)
1=cap diameter	12.70
2=cavity diameter	9.65
3=apex diameter	2.54
4=cavity depth	0.53
5=cap thickness	0.25
6=stud height	6.35
7=stud diameter	1.45

B. Projectors

The projectors compared in this paper consist of 25 "cymbal" electromechanical drivers sandwiched between and mechanically fastened to two lightweight, stiff cover plates with dimensions 7.6 cm × 7.6 cm × 0.23 cm. The cover plates are made from a commercially available graphite-epoxy composite.[23] After the plates are electro-

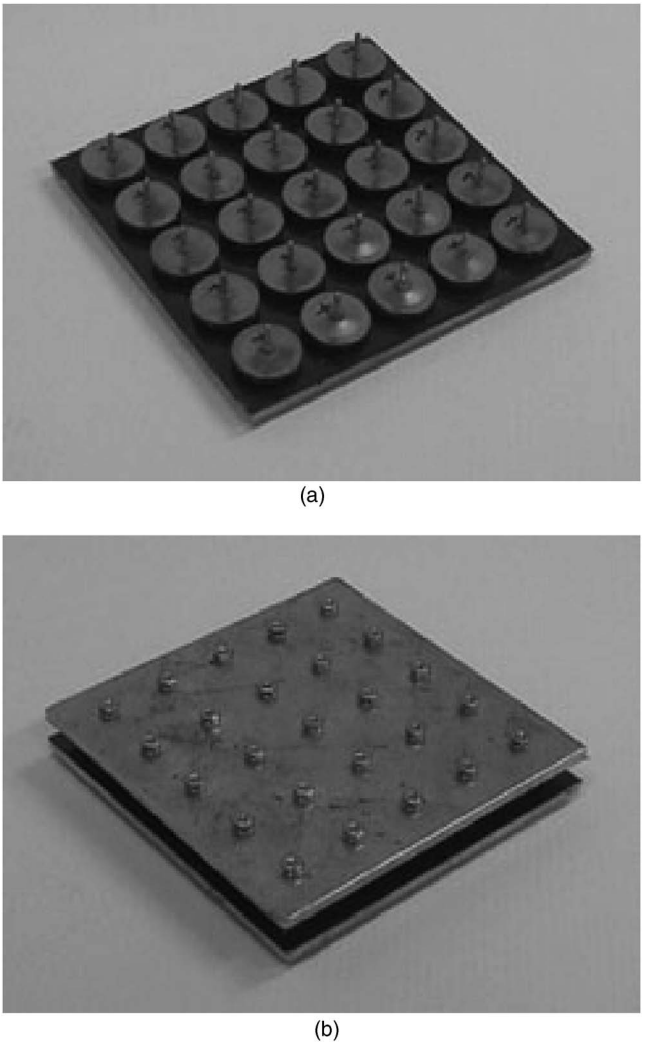


FIG. 4. Cymbal projector showing (a) cymbal elements inside and (b) with top cover plate attached.

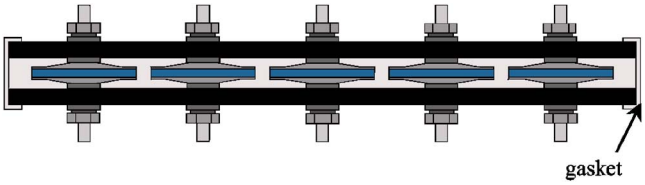


FIG. 5. (Color online) Side-view cutaway sketch of the cymbal projector showing the outer gasket in place.

plated with copper, holes that will allow the cymbal stud to pass through are drilled. Holes are drilled through two plates at once (making sure that the electroplated surfaces are facing outward) to make a matched pair of plates. Electrically insulating Kapton® polyimide adhesive tape is applied to the "inside" surface of each of the cover plates and holes are poked through the tape to open up the through holes.

The cymbal elements are fastened to the cover plates using hex nuts, flat washers, and lock washers as shown in Fig. 4. Leads from the measurement cable are soldered to the electroplated surfaces of the projector module panel. Prior to this, a polyurethane gasket is wrapped around the outside edge of the module in order to seal in the air pocket between the cover plates (Fig. 5). It is necessary to maintain this air gap between the cover plates in order to produce the low frequency resonance behavior. Furthermore, if urethane leaks into this space during the potting process, the flexural motion of the cymbal caps is compromised resulting in a reduced acoustic output and directivity.

The final stage in the cymbal panel fabrication process is encapsulation of the projector with polyurethane. Because of the need to retain the encapsulated air matrix between the cover plates within the projector, it is not practical to pull a vacuum on the polyurethane during the curing stage. Instead, the polyurethane is heated in an oven and, in order to minimize the formation of air bubbles within the urethane, a vacuum is pulled on it prior to its being introduced into the overmold. A finished cymbal module projector is shown in Fig. 6.

III. IN-AIR CHARACTERIZATION OF CYMBAL ELEMENTS

Prior to being incorporated into the projectors, the individual cymbal elements were characterized in air. The data are summarized in Table III. The 1 kHz capacitance and dielectric loss of the piezoelectric ceramic disks were very consistent, i.e., the standard deviation for the capacitance was $<8\%$. The single crystal disks, however, exhibited a much wider variation in capacitance, showing a standard deviation of 34%.

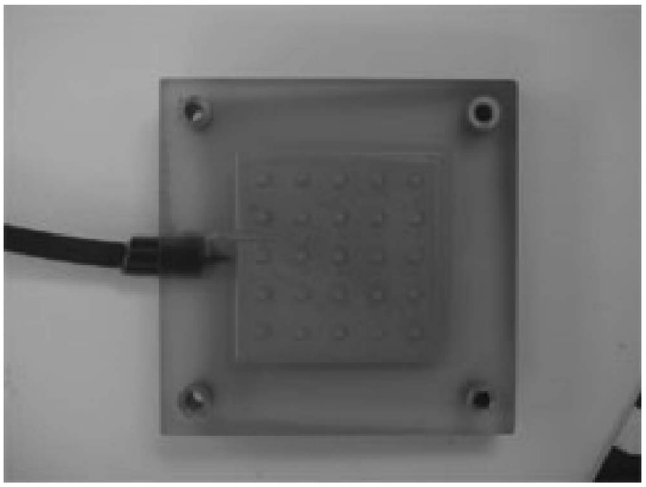


FIG. 6. Cymbal projector potted in polyurethane with measurement cable attached.

When the disks were incorporated into cymbal elements, a drop in capacitance was seen. This reduction in capacitance was especially pronounced (down by $\sim 45\%$) in the single crystal-based cymbals and is attributed to lateral clamping of the elastically soft crystal by the stiffer titanium caps.[24,25] Note that the Young's modulus (Y_{11}) of the single-crystal material is at least 70% lower than that of a PZT ceramic (Table I). The resonance frequencies of the individual cymbal elements are also consistent with the stiffness of the piezoelectric material and the variation ($0.40 < \sigma_{n-1} < 0.46$) is thus essentially independent of material type.

IV. UNDERWATER CALIBRATION

A. Experimental setup and facilities

The projectors were first calibrated in August 2003 at the Naval Sea Systems Command, Division Crane, Glendora Lake Facility[26] in Sullivan, Indiana. The lake temperature was 7.6 °C. Measurements were taken with the projector held at its two top corners with nylon wire wraps. A receiver hydrophone was positioned at the same depth as the projec-

TABLE III. Single element cymbal characteristics.[a]

	Piezoelectric disk only (measured at 1 kHz)		Cymbal (measured at 1 kHz)			
Projector type	C (nF)	tan δ	C (nF)	tan δ	f_r (kHz)	Adhesive pull strength (kg)
PZT-4	1.46	0.0036	1.36	0.003	31.63	13.3±0.3
	$\sigma_{n-1}=0.03$	$\sigma_{n-1}=0.0006$			$\sigma_{n-1}=0.40$	
PZT-5H	3.22	0.014	3.02	0.014	31.54	
	$\sigma_{n-1}=0.08$	$\sigma_{n-1}=0.001$			$\sigma_{n-1}=0.46$	
PMN-30%PT	6.40	0.0033	3.54	0.0046	28.09	7.5±0.6
	$\sigma_{n-1}=0.34$	$\sigma_{n-1}=0.0003$	$\sigma_{n-1}=0.17$	$\sigma_{n-1}=0.0007$	$\sigma_{n-1}=0.40$	

[a]Standard deviation is reported when multiple samples were available for measurement within each group.

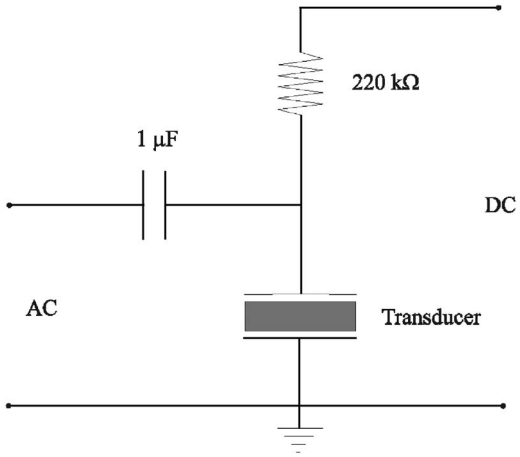


FIG. 7. Diagram of the dc bias circuit used for the single crystal cymbal projector.

tor (12.2 m) with a separation distance between the two of 5.16 m. The applied electrical drive signal was a normal cw pulse with duration 8 ms (0.5 to 8 kHz sweep) or 3 ms (5 to 25 kHz sweep) and a 1% duty cycle. The receive signal was gated to eliminate interference from reflections off the surface of the water. The results from this initial calibration are presented in Sec. IV B 1 and IV B 2.

The projectors were recalibrated in March 2004 at the Acoustic Test Facility,[27] Naval Sea Systems Command, in Newport, Rhode Island. The size of their test tank is 18 m $\times$12 m$\times$11.5 m and no two sides are parallel. The water temperature was 16.1 °C. The separation distance between the test projector and the standard H-52 hydrophone was 2 m and the test depth was varied from 5.2 to 7.5 m. The applied drive signal was a 30 ms cw tone at 3 pulse/s (10% duty cycle). The top of the projectors were hard mounted to a test fixture. It was determined experimentally that hard mounting the projector to a test fixture gave the same results as when the projector was held at its two top corners with parachute cord. The data from this measurement set are summarized in Sec. IV B 3.

Because of the very low coercive field associated with single-crystal piezoelectric materials (see Table I), a dc bias voltage needs to be applied in order to take full advantage of the high strain associated with these materials. We tried to maintain a dc bias level of at least $\sqrt{2}$ times the ac peak drive level. The dc bias circuit used for these measurements is diagrammed in Fig. 7.

TABLE IV. Drive conditions for the single crystal-based projector.

Key corresponding to Figs. 8, 9, and 12	Key corresponding to Fig. 11
A=100 V_{rms}+165 V_{dc}	*A*=102 V_{rms}+165 V_{dc}
B=150 V_{rms}+220 V_{dc}	*B*=148 V_{rms}+220 V_{dc}
C=200 V_{rms}+300 V_{dc}	*C*=194 V_{rms}+300 V_{dc}
D=250 V_{rms}+365 V_{dc}	*D*=247 V_{rms}+365 V_{dc}
E=300 V_{rms}+442 V_{dc}	*E*=290 V_{rms}+442 V_{dc}
F=350 V_{rms}+499 V_{dc}	*F*=341 V_{rms}+499 V_{dc}
G=380 V_{rms}+622 V_{dc}	*G*=369 V_{rms}+622 V_{dc}

B. Results and discussion

The projector data presented in this section have been corrected to remove the effects of the measurement cable.

1. Single crystal projector

Figure 8 shows the acoustic source level generated by the single crystal-based projector as a function of ac drive field and associated dc bias level. The drive conditions, given in the left half of Table IV, were the maximum drive conditions seen by the projector. Figure 9 shows the actual ac voltage applied to the projector as a function of frequency. Because the impedance of the projector changes across the frequency band, the ac drive level is not constant. The associated dc bias voltage given in Table IV was constant across the entire band from 500 Hz–25 kHz. The highest field applied to the projector was 380 V_{rms}+622 V_{dc}. This was the highest that we could go as far as the dc power supply was concerned. The projector, in fact, showed no signs of degraded performance when the source level was subsequently rerun at 100 V_{rms}+165 V_{dc}. Figure 10 shows the acoustic source level generated by the projector driven at its in-water

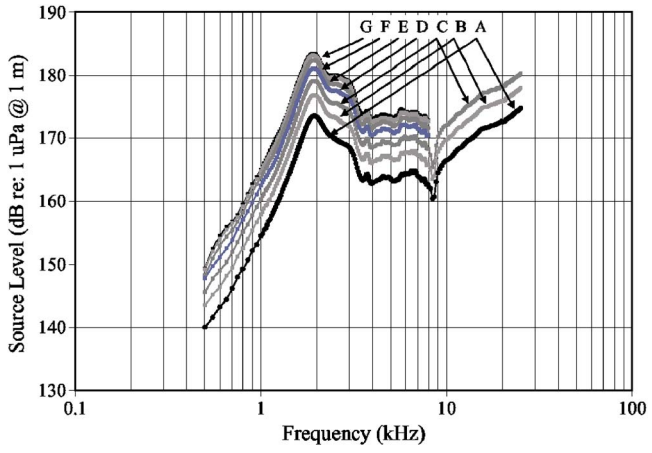


FIG. 8. (Color online) Acoustic source level measured from the single crystal cymbal projector as a function of frequency and drive level. Maximum drive conditions are given in the left half of Table IV.

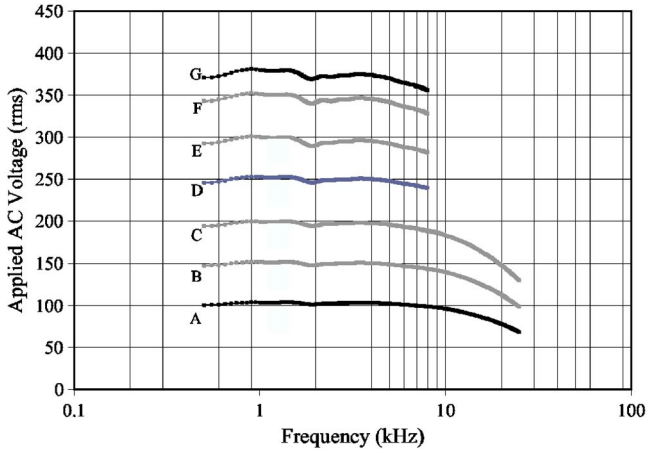


FIG. 9. (Color online) Actual ac voltage applied to the single crystal cymbal projector as a function of frequency. The letters to the left of these curves are used in Figs. 8, 11, and 12 for simplicity. The dc voltages shown in Table IV are constant with frequency.

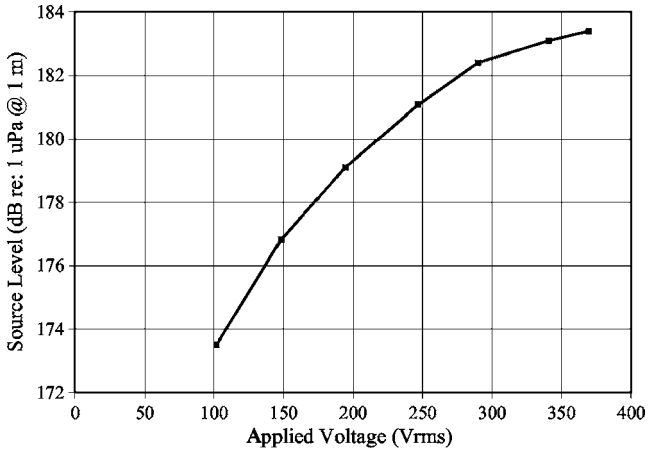


FIG. 10. Acoustic source level generated by the single crystal cymbal projector at its in-water resonance frequency as a function of applied ac voltage.

resonance frequency of 1.94 kHz as a function of applied ac voltage (and appropriate dc bias). The source level is, in fact, not linear with drive field but tails off when the ac drive level exceeds $\sim$300 V_{rms}.

Looking at the transmitting voltage response (TVR) in the neighborhood of the projector in-water resonance frequency, Fig. 11, it is observed that the TVR begins a significant drop when the drive level exceeds $\sim$300 V_{rms}. The right half of Table IV shows the drive conditions applied to the projector at its resonance frequency of 1.94 kHz. Based on this data, we contend that the maximum safe drive level applied to the single-crystal-based cymbal module is 300 V_{rms}+442 V_{dc}.

The electrical impedance and phase angle as a function of drive level are compared in Fig. 12. The data show that the impedance is independent of drive level from 500 Hz to 25 kHz. The phase angle remains unchanged as a function of drive level until $\sim$300 V_{rms}+442 V_{dc} is reached. This is in agreement with the TVR data of Fig. 11, confirming that the maximum drive level for this projector should not exceed 300 V_{rms}+442 V_{dc}.

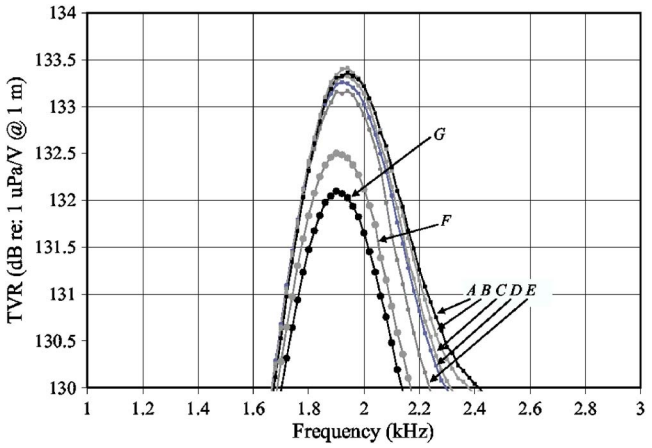


FIG. 11. (Color online) Transmitting voltage response of the single crystal cymbal projector as a function of drive level in the neighborhood of its in-water resonance frequency. The drive conditions given in the right half of Table IV are at the resonance frequency.

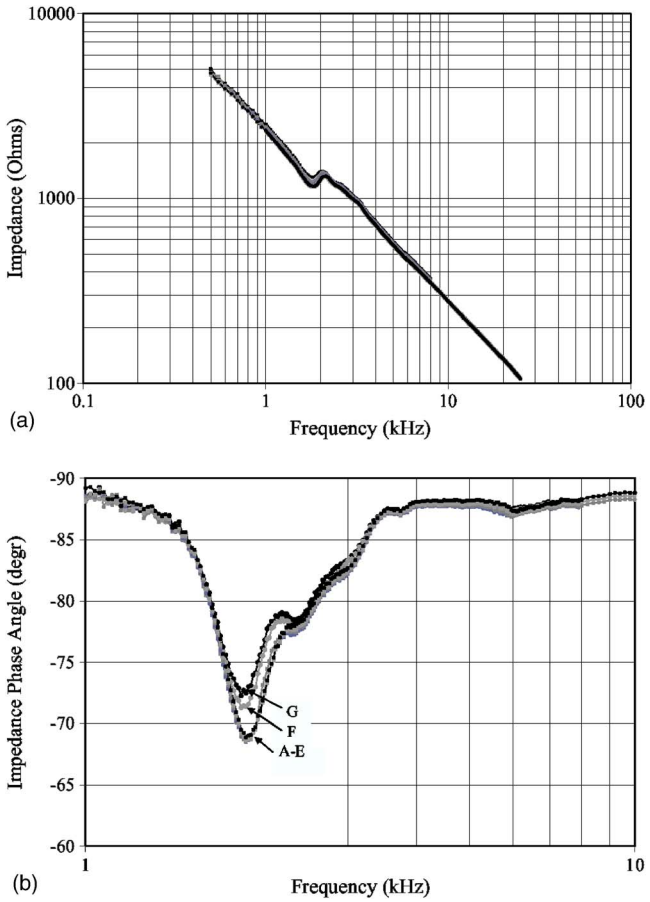


FIG. 12. (Color online) (a) Impedance and (b) phase angle of the single crystal cymbal projector as a function of frequency and drive level. Note that the frequency scales between (a) and (b) are different. See Table IV and Fig. 9 for drive conditions.

2. Single-crystal-versus PZT-based projectors

Figure 13 compares the transmitting voltage response (TVR) measured for the PZT-4-, PZT-5H-, and single crystal-based projectors. All three projectors behave as expected, with the single crystal-based unit exhibiting $\sim$5 dB improvement in response compared with the PZT-5H unit and nominally an 8 dB enhancement over the PZT-4-based projector. The single crystal-based unit also shows a slightly

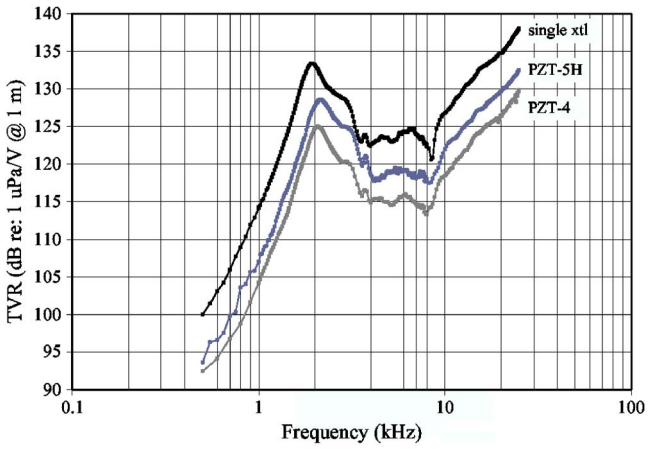


FIG. 13. (Color online) Comparison of the transmitting voltage responses of the PZT-4-, PZT-5H-, and single crystal-based cymbal projectors.

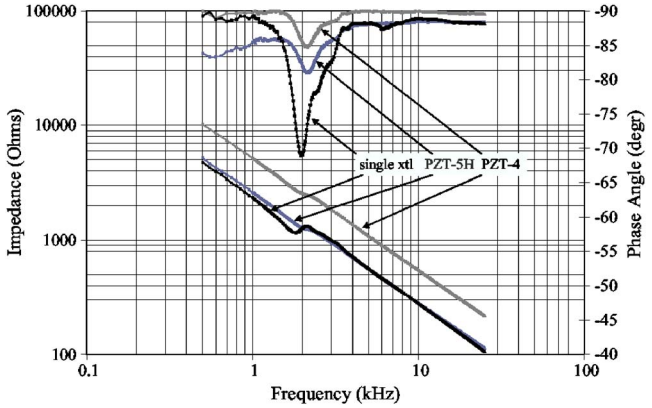


FIG. 14. (Color online) Comparison of the impedance and phase angle of the PZT-4-, PZT-5H-, and single crystal-based cymbal projectors.

lower resonance frequency as compared to the PZT-based units because of the slightly lower mechanical stiffness associated with single crystal material.

The electrical impedance and phase angle of the three projectors are compared in Fig. 14. The impedance magnitude is consistent with the cymbal capacitance values measured at 1 kHz (Table III). Because the capacitance of the single crystal cymbals and PZT-5H cymbals are nearly the same, their impedance magnitudes are quite similar. The PZT-4 cymbals have a significantly lower capacitance, so their impedance is higher. The most notable difference between the three projectors is in the phase angle, with the single crystal projector having the broadest and most pronounced of the three.

The source levels for the three cymbal panels, based on their maximum recommended drive levels, are compared in Fig. 15. The single crystal-based unit exhibits the highest source level across the band of interest—approximately 4 dB higher than the PZT-4-based unit and 9 dB higher than the PZT-5H-based unit. Even though the PZT-4-based cymbal panel has a lower TVR than the PZT-5H-based projector, its maximum source level is higher because it has a higher electrical impedance and can be safely driven with a higher voltage.

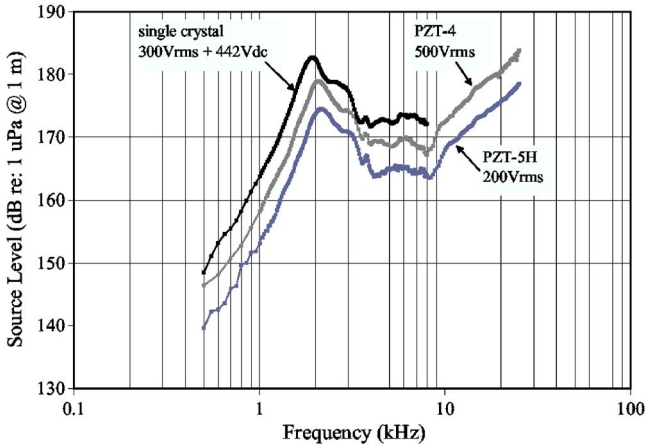


FIG. 15. (Color online) Comparison of the acoustic source level obtained from the PZT-4-, PZT-5H-, and single crystal-based cymbal projectors when driven at their maximum safe drive levels.

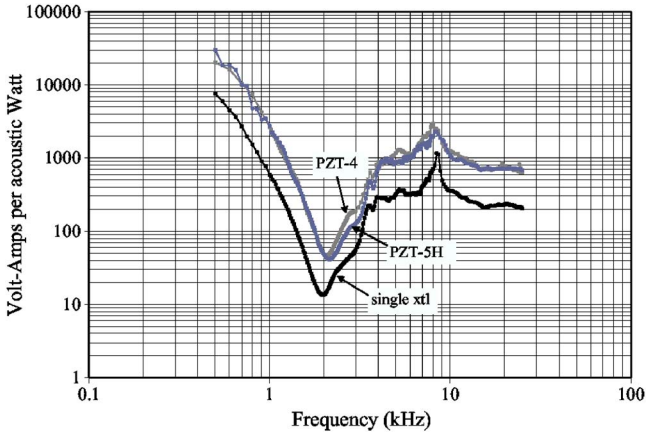


FIG. 16. (Color online) Comparison of the volt-amp per acoustic watt performance of the PZT-4-, PZT-5H-, and single crystal-based cymbal projectors.

Perhaps the most objective parameter that can be used to compare the three projectors is the volt-amp product required to achieve 1 W of acoustic output. This is shown in Fig. 16. The two PZT-based projectors are practically the same over the frequency band of interest. The single crystal-based projector, on the other hand, requires nominally one-third less power than a PZT-based unit to produce 1 W of acoustic energy.

Other parameters often used to compare underwater acoustic projectors are the mechanical Q_M, electrical Q_E, effective coupling coefficient k_{eff}, and electroacoustic efficiency η. These parameters are provided in Table V. The mechanical Q_M and Q_E are shown for the modules both in air and in water. They were calculated from Eqs. (3) and (4) below,[28] where f_S is the series resonance frequency, f_a is the frequency of maximum resistance, f_2 and f_1 are the frequencies at which the motional contribution to the conductance is half its maximum value, and G_{mS} and B_S are the respective conductance and susceptance values measured at the resonance frequency.

$$Q_M = \frac{f_S}{f_2 - f_1}, \tag{3}$$

$$Q_E = \frac{B_S}{G_{mS}}. \tag{4}$$

The effective coupling coefficient was subsequently calculated from the in-air Q_M and Q_E values as[29]

$$k_{\text{eff}} = \sqrt{\frac{1}{1 + Q_M Q_E}} \quad \text{when } 10 < Q_M < 50, \tag{5a}$$

$$k_{\text{eff}} = \sqrt{1 - \left(\frac{f_S}{f_a}\right)^2} \quad \text{when } Q_M > 50. \tag{5b}$$

The electroacoustic efficiency value given in Table V is that calculated for the in-water resonance frequency. It was determined by solving Eq. (6), where the directivity index

TABLE V. Additional characteristics of the cymbal projectors.

	In air				In water			
Projector type	f_s	Q_M	Q_E	k_{eff}	f_s	Q_M	Q_E	η at f_r
PZT-4	6.6	49	0.77	0.16	2.1	3	11	~22%
PZT-5H	6.6	36	0.92	0.17	2.2	1.5–2.5	6	~16%
PMN-30%PT	5.8	53	0.095	0.31	2.0	2–3	3	~21%

(DI) was assumed to be the same as that of an unbaffled circular piston[30] with an equivalent aperture size as the cymbal module.

$$\eta = 10^{(\text{TVR}-10\log G-170.8-\text{DI})/10}. \tag{6}$$

The in-air data from Table V show that the k_{eff} of the single crystal projector is nominally twice that of the PZT-based units. This is due to the higher coupling associated with the single crystal material. The in-water electroacoustic efficiencies are all about 20% and are consistent with the respective in-water Q_M values, to which η is directly proportional. The low radiation efficiency is a consequence of the projector being small compared to the wavelength in water around resonance.

Figure 17 shows the directivity patterns measured for the three projectors at 25 kHz. The theoretical pattern for a circular piston[30] is superposed on the PZT-4 plot. Both the PZT-4 and PZT-5H units are seen to correlate quite well with the theoretical prediction. However, the single crystal-based unit exhibits slight asymmetry, which is attributed to greater nonuniformity in piezoelectric properties among the single crystal elements. Because of the wider variation in capacitance (and presumably d_{31} as well), the displacement of each single crystal cymbal will vary, and hence the motion at the top of the cover plate will be nonuniform, giving rise to the asymmetric beam pattern. Furthermore, the effect of the manually applied adhesive on the single crystal elements can cause considerable variations within each element and from element to element as well.

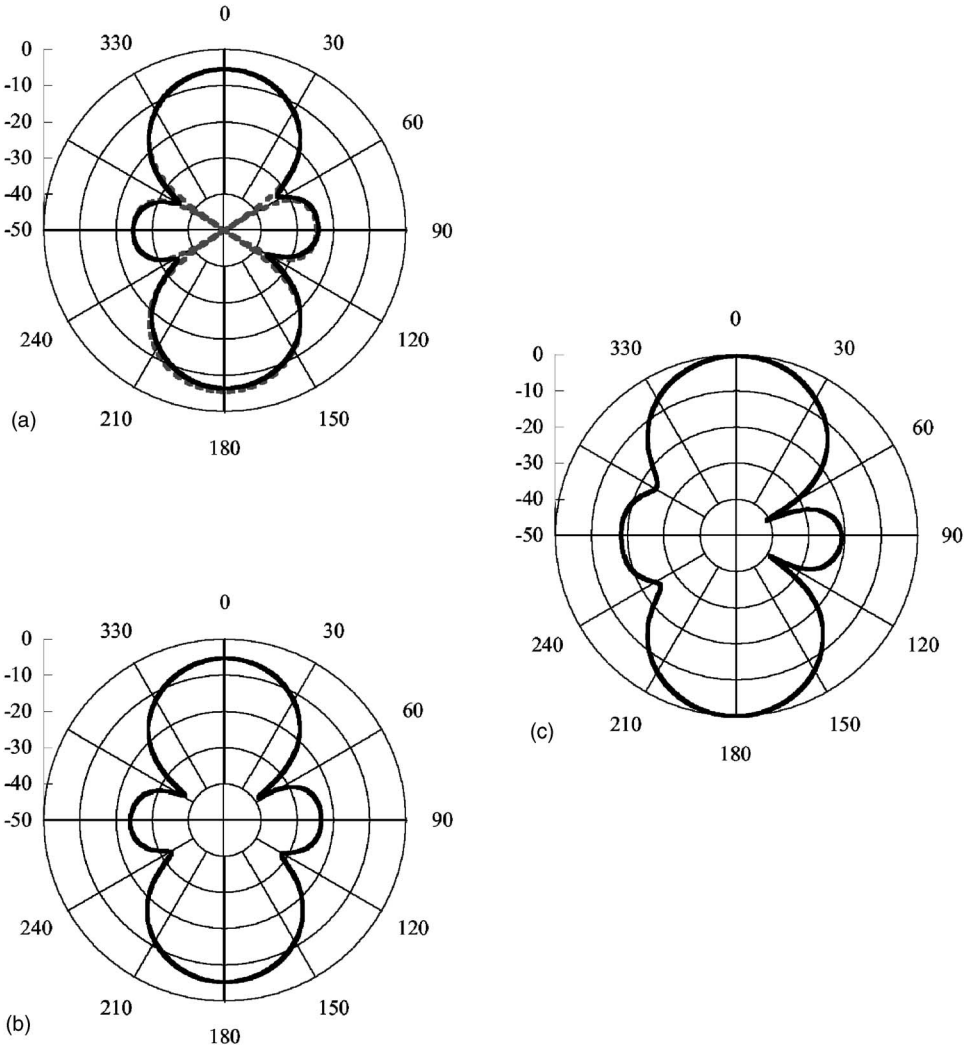


FIG. 17. Directivity patterns measured at 25 kHz for the (a) PZT-4-, (b) PZT-5H-, and (c) single crystal-based cymbal projectors. The dashed curve in (a) shows the theoretical prediction.

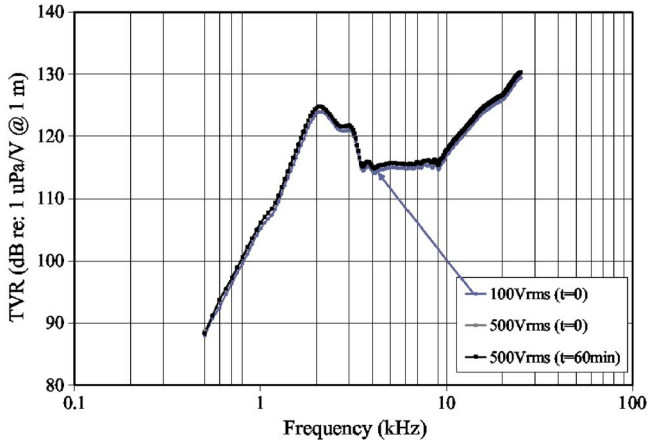


FIG. 18. (Color online) Comparison of the transmitting voltage response from the PZT-4-based cymbal projector when driven at 100 and 500 V_{rms} (both initially and after 1 h).

3. Performance at high drive levels and fatigue tests

This set of measurements was performed at the NUWC-Newport facility in March 2004. The projectors were all driven at their maximum recommended drive level (see Fig. 15) using a pulsed cw signal with a 10% duty cycle. The purpose was to determine whether the acoustic performance of the projectors degraded over time. Figure 18 compares the TVR measured for the PZT-4-based projector driven at 100 V_{rms} to that at 500 V_{rms}. No significant difference in performance is seen. Likewise, the TVR measured for the projector after it was run continuously at its resonance frequency for 60 min at 500 V_{rms} is identical to that measured initially.

Figure 19 shows that the TVR for the PZT-5H-based projector remains essentially unchanged when going from a 100 to 200 V_{rms} drive level. This projector was subsequently run at 200 V_{rms} for ten minutes at its resonance frequency before stopping to perform a TVR sweep from 500 Hz to 25 kHz. After repeating this process for two hours, no change was seen in the transmitting response.

As mentioned in Sec. IV A, all three projectors were initially calibrated in August 2003. After this set of measurements, the projectors sat dormant in air for nearly seven months before being recalibrated at the NUWC-Newport facility in March 2004. The transmitting voltage response curves for the three projectors for the two measurement times are compared in Fig. 20. The before-and-after TVRs from both the PZT-4- and PZT-5H-based projectors are within 0.5 dB (i.e., measurement error) of each other, respectively. However, the TVR of the single crystal-based projector had dropped by ~6 dB at its resonance frequency. As the capacitance of the unit did not change over this time, the decrease in TVR is attributed to degradation in the cap-to-crystal bonding layer.

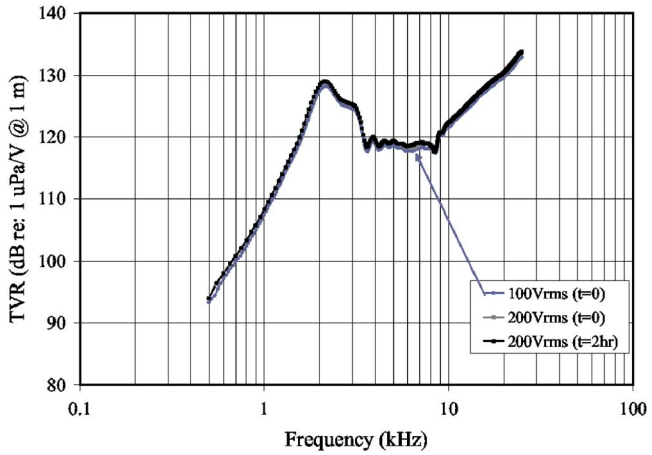


FIG. 19. (Color online) Comparison of the transmitting voltage response from the PZT-5H-based cymbal projector when driven at 100 and 200 V_{rms} (both initially and after 2 h).

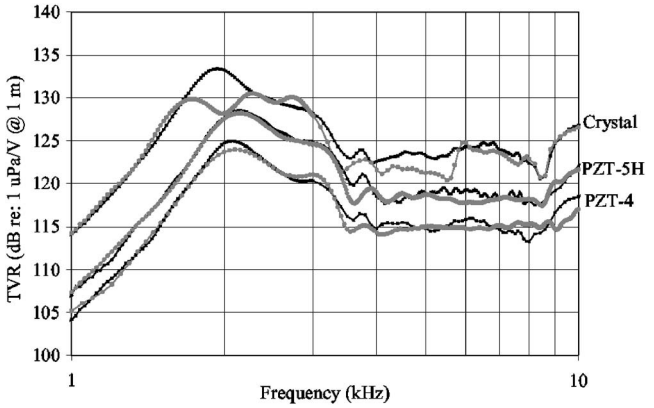


FIG. 20. Comparison of the transmitting voltage response of the PZT-4, PZT-5H, and single crystal cymbal projectors measured in August 2003 (black) and March 2004 (gray).

The strength of the adhesive in an individual cymbal element was measured by a simple pull test. The top cymbal stud was fastened to a jig that maintained the vertical alignment of the studs. A fixture to which mass could be added was screwed into the lower stud. Mass was then continually added to this fixture until the bond broke.

The adhesive used to make the cymbals in the single crystal-based unit was not as strong as that used for the PZT-based units (see Table III). Therefore, over time the mechanical coupling between the radial motion in the single crystal disk and the flexural motion of the caps degraded. This weakening of the bond would be most evident at the resonance frequency. The lower strength of this two-part epoxy adhesive as compared to the one component adhesive film is likely due to several factors. One is simply differences in chemical composition. Another is the presence of air bubbles in the two-part epoxy mixture that formed during the mixing process which act to weaken the bond. A third factor is the smaller thickness of the epoxy layer (~25 μm) as compared to that of the adhesive film (50 μm). Single crystal cymbals with stronger crystal-to-cap bonds have subsequently been fabricated using a one component low temperature cure adhesive film similar to that used for the PZT-based cymbals.[25]

Figure 21 compares the TVR measured initially at a drive level of 300 V_{rms}+442 V_{dc} to that measured after two hours, with the experiment being run the same as the PZT-5H fatigue experiment described previously. There is a slight reduction in performance in the neighborhood of the resonance as well as in the band 4–8 kHz. This is attributed to further degradation of the bonding layer.

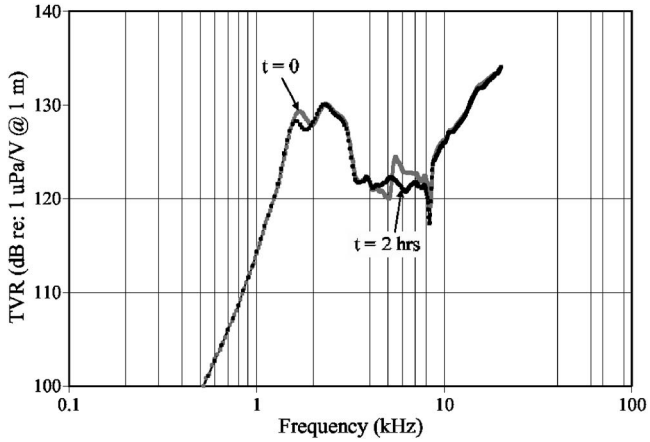


FIG. 21. Comparison of the transmitting voltage response from the single-crystal based cymbal projector when driven at 300 V_{rms}+442 V_{dc} (both initially and after 2 h).

V. CONCLUSIONS

The cymbal panel underwater electroacoustic projector that utilized single crystal relaxor ferroelectrics with composition PMN-30%PT can show at least a 4 dB improvement in acoustic source level as compared to identical units made from PZT piezoelectric ceramic. Most notably, the single crystal-based projector requires one third less power to generate 1 W of acoustic energy from 500 Hz to 25 kHz. However, since we were operating well below the fundamental resonance frequency of the single crystal material, we did not take advantage of its large electromechanical coupling feature and the wider operating bandwidth associated with it. In essence, we really only utilized the superior piezoelectric and high drive capabilities of the single crystal material. No degradation in acoustic performance was seen in the PZT-based projectors when driven under high electric field for lengthy times. Some slight reduction was observed in the single crystal-based projector, but this is believed to be due to degradation in the bonding layer rather than from the crystals themselves.

In future design iterations of the cymbal panel, it will be necessary to utilize a better cap-to-crystal bonding methodology. In addition, property variation amongst the individual crystals must be improved by the commercial vendors, primarily to eliminate asymmetry in the directivity patterns. Last, it would be prudent to consider a redesign of the cymbal elements so as to take advantage of the extremely high electromechanical coupling capability of the single crystals.

ACKNOWLEDGMENTS

The authors would like to thank DARPA and ONR for their support of this project. They would also like to thank TRS Technologies in State College, PA for providing the single crystals. Additional thanks to Walter Carney and Scott Small of Navsea Division Crane, Kirk Robinson and Mel Jackaway of Navsea Division Crane, Glendora Lake facility, and Walter Boober and Rene LeFleur of Navsea Division Newport.

[1]S.-E. Park and W. Hackenberger, "High performance single crystal piezoelectrics: Applications and issues," Curr. Opin. Solid State Mater. Sci. **6**, 11–18 (2002).
[2]T. Ritter, X. Geng, K. K. Shung, P. D. Lopath, S.-E. Park, and T. R. Shrout, "Single crystal PZN/PT-polymer composites for ultrasound transducer applications," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 792–800 (2000).
[3]C. G. Oakley and M. J. Zipparo, "Single crystal piezoelectrics: A revolutionary development for transducers," *2000 IEEE Ultrasonics Symposium* (IEEE Service Center, Piscataway, NJ, 2000), pp. 1157–1167.
[4]P. W. Rehrig, W. S. Hackenberger, X. Jiang, R. J. Meyer, and X. Geng, "Naval device applications of relaxor piezoelectric single crystals," *2002 IEEE Ultrasonics Symposium* (IEEE Service Center, Piscataway, NJ, 2002), pp. 733–737.
[5]E. A. McLaughlin and H. C. Robinson, "Single-crystal ferroelectrics for sonar devices," J. Acoust. Soc. Am. **116**, 2590 (2004).
[6]S.-E. Park and T. R. Shrout, "Ultrahigh strain and piezoelectric behavior in relaxor based ferroelectric single crystals," J. Appl. Phys. **82**, 1804–1811 (1997).
[7]L. E. Cross, "Relaxor ferroelectrics," Ferroelectrics **76**, 241–267 (1987).
[8]Channel Industries, Inc., *Piezoelectric Ceramics*, product catalog provided by Channel Industries, Inc., Santa Barbara, CA, 2004.
[9]J. Yin, B. Jiang, and W. Cao, "Elastic, piezoelectric, and dielectric properties of $0.955Pb(Zn_{1/3}Nb_{2/3})O_3-0.45PbTiO_3$ single crystal with designed multidomains," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 285–291 (2000).
[10]R. Zhang, B. Jiang, and W. Cao, "Elastic, piezoelectric, and dielectric properties of multidomain $0.67Pb(Mg_{1/3}Nb_{2/3})O_3-0.33PbTiO_3$ single crystals," J. Appl. Phys. **90**, 3471–3475 (2001).
[11]R. Zhang, W. Jiang, B. Jiang, and W. Cao, "Elastic, dielectric and piezoelectric coefficients of domain engineered $0.70Pb(Mg_{1/3}Nb_{2/3})O_3-0.30PbTiO_3$ single crystal," in *Fundamental Physics of Ferroelectrics 2002, AIP Conference Proceedings Vol 626*, edited by R. E. Cohen (American Institute of Physics, Melville, NY, 2002), pp. 188–197.
[12]W. P. Mason, "Piezoelectricity, its history and applications," J. Acoust. Soc. Am. **70**, 1561–1566 (1981).
[13]H. Jaffe and D. A. Berlincourt, "Piezoelectric transducer materials," Proc. IEEE **53**, 1372–1386 (1965).
[14]O. B. Wilson, *Introduction to Theory and Design of Sonar Transducers* (Peninsula Publishing, Los Altos, CA, 1988), Chap. 4, pp. 65–88.
[15]S.-E. Park and T. R. Shrout, "Characteristics of relaxor-based piezoelectric single crystals for ultrasonic transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1140–1147 (1997).
[16]T. R. Shrout, R. Eitel, S. Zhang, C. Randall, E. Alberta, and P. Rehrig, "Recent developments in transition temperature (T_c) Perovskite crystals," *2003 IEEE Ultrasonics Symposium* (IEEE Service Center, Piscataway, NJ, 2003), pp. 774–777.
[17]R. J. Meyer, Jr., W. J. Hughes, and T. R. Montgomery, "Single crystal tonpilz transducers," presented at *2003 U.S. Navy Workshop on Acoustic Transduction Materials and Devices* (Materials Research Institute, The Pennsylvania State University, University Park, PA, 2003).
[18]T. R. Howarth, K. C. Benjamin, D. Huang, and J. F. Tressler, "Broadband, lower-frequency acoustic projector design for AUV applications," J. Acoust. Soc. Am. **115**, 2619 (2004).
[19]R. E. Newnham and A. Dogan, "Metal-electroactive ceramic composite transducer," U.S. Patent No. 5,729,077, 1998.
[20]R. A. Nelson, Jr. and L. H. Royster, "Development of a mathematical model for the class V flextensional underwater acoustic transducer," J. Acoust. Soc. Am. **49**, 1609–1620 (1971).
[21]A. Dogan, K. Uchino, and R. E. Newnham, "Composite piezoelectric transducer with truncated conical endcaps 'cymbal'," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 597–605 (1997).
[22]J. F. Tressler and T. R. Howarth, "Cymbal drivers utilizing relaxor-based ferroelectric single crystal materials," *Proceedings of the 2000 12th IEEE International Symposium on Applications of Ferroelectrics* (IEEE Service Center, Piscataway, NJ, 2001), pp. 561–564.
[23]Aerospace Composite Products, San Leandro, CA, www.acp-composites.com (November 2005).
[24]Q.-M. Wang, X.-H. Du, B. Xu, and L. E. Cross, "Theoretical analysis of the sensor effect of cantilever piezoelectric benders," J. Appl. Phys. **85**, 1702–1712 (1999).
[25]J. F. Tressler, "Fabrication and characterization of single crystal cymbals," Ferroelectrics (to be published).

[26]http://www.crane.navy.mil/acoustics/glendoralake/glendora.asp (November 2005).
[27]http://www.npt.nuwc.navy.mil/ATF8211/index.html (November 2005).
[28]D. Stansfield, *Underwater Electroacoustic Transducers* (Bath University Press, Bath, UK, 1991), Chap. 4–3, pp. 78–83.
[29]M. B. Moffett and W. J. Marshall, Jr., "The importance of coupling factor for underwater acoustic projectors," NUWC-NPT Technical Document 10,691, 1994.
[30]L. L. Beranek, *Acoustics* (Acoustical Society of America through the American Institute of Physics, Woodbury, NY, 1993), Chap. 4, pp. 91–115.

Optimally sensitive and efficient compact loudspeakers

Ronald M. Aarts
Philips Research, High Tech Campus 36, NL-5656 AE Eindhoven, The Netherlands

(Received 3 August 2005; revised 7 November 2005; accepted 18 November 2005)

In conventional loudspeaker system design, the force factor *Bl* is chosen in relation to enclosure volume, cone diameter, and moving mass to yield a flat response over a specified frequency range. For small-cabinet loudspeakers such a design is quite inefficient. This is shown by calculating the efficiency and voltage sensitivity. The frequency response is manipulated electronically in a strong nonlinear fashion, which has consequences for the sound quality, but it then turns out that systems using much lower force factors can provide greater usable efficiency, at least over a limited frequency range. For these low-force-factor loudspeakers, a practically relevant and analytically tractable optimality criterion, involving the loudspeaker parameters, will be defined. This can be especially valuable in designing very compact loudspeaker systems. An experimental example of such a design is described. This new, optimal design has a much higher power efficiency as well as a higher voltage sensitivity than current bass drivers, while the cabinet can be much smaller. [DOI: 10.1121/1.2151694]

PACS number(s): 43.38.Ja, 43.38.Dv [AJZ] Pages: 890–896

I. INTRODUCTION

There is a longstanding interest in obtaining a high sound output from compact loudspeaker arrangements. Compact relates here to both the volume of the cabinet in which the loudspeaker is mounted as well as to the cone area of the loudspeaker. Loudspeakers can be built such that they properly reproduce the entire audible frequency spectrum, down to 20 Hz; but such systems would be both expensive and very bulky. In many sound reproduction applications it is not possible to use large loudspeaker systems because of size or cost constraints. Typical applications are portable audio, multimedia, and (flat) TV sets. Various signal-processing schemes have been proposed to equalize the response of small loudspeakers or to use psychoacoustic enhancement methods; see Larsen and Aarts (2004) for some overview. The aim of this present paper is to discuss a method to manipulate electronically, in a strong nonlinear fashion, a special loudspeaker with a high acoustical output. The dependence of the transducer's behavior on various parameters, in particular the force factor *Bl*, is investigated. For electrodynamic loudspeakers the perceived quality is important, but also the sensitivity [Pa/V] and the efficiency are of importance. Therefore, in the following section the sensitivity and the efficiency of electrodynamic loudspeakers in general are discussed. It appears that drivers with very high efficiency have poor sensitivity at low frequencies. It is not possible to combine a very high efficiency and a high sensitivity in a wide frequency range with a compact arrangement. In Sec. III special drivers with a very low—but optimal—*Bl* value will be discussed. They have an optimal sensitivity and are only 3 dB less efficient than an infinite-force-factor loudspeaker, but in a limited frequency range only. These characteristics are obtained at the expense of decreased sound quality and the requirement of some additional electronics. Due to the low-*Bl* value, the magnet can be considerably smaller than usual and the loudspeaker can be of the moving-magnet type with a stationary coil, instead of vice versa. In Sec. IV it is discussed how such a low-*Bl* driver can be made. It appears to be very cost-efficient, low-weight, flat, and requires a low-volume cabinet.

II. SENSITIVITY AND EFFICIENCY CALCULATIONS

For low frequencies a loudspeaker can be modeled using some simple elements, allowing the formulation of approximate analytical expressions for the loudspeaker sound radiation (Beranek, 1954; Thiele, 1971; Small, 1972). Neglecting the self-inductance L_e of the driver's voice coil, the transfer function at distance r from voltage $E(s)$ to pressure $P(s)$, also known as the sensitivity, can be written (Aarts, 2005) as

$$H_p(s) = \frac{P(s)}{E(s)} = \frac{s^2 \rho S/(2\pi r) Bl/R_e}{s^2 m_t + sR_t + k_t}, \tag{1}$$

with all used variables as listed in Table I. It appears to be convenient to use the following dimensionless quality factors Q, the dimensionless frequency detuning ν, and resonance frequency ω_0

$$Q_m = \sqrt{k_t m_t}/R_m, \quad Q_e = R_e \sqrt{k_t m_t}/(Bl)^2,$$

$$Q_r = \sqrt{k_t m_t}/R_r, \quad Q_{mr} = Q_m Q_r/(Q_m + Q_r),$$

$$Q_t = (m_t \omega_0)/R_t, \quad \nu = \omega/\omega_0 - \omega_0/\omega, \text{ and } \omega_0 = \sqrt{k_t/m_t}. \tag{2}$$

Using these equations, Eq. (1) can be written as

$$H_p(\omega) = \frac{(i\omega/\omega_0)^2}{(i\omega/\omega_0)^2 + (i\omega/\omega_0)Q_t^{-1} + 1} \left(\frac{\rho a^2 Bl}{2 m_t r R_e} \right). \tag{3}$$

The first fraction on the right-hand side of Eq. (3) expresses the typical high-pass characteristic of a loudspeaker, while the second fraction gives the value for high frequencies ($\omega \gg \omega_0$). At the resonance frequency ($\omega = \omega_0$) Eq. (3) becomes

TABLE I. System parameters of the model.

a	radius of the cone
B	flux density in the air gap
Bl	force factor
E	voice coil voltage
F	$=BlI_c$ is the Lorentz force acting on the voice coil
i	$\sqrt{-1}$
I_c	voice coil current
k_t	total spring constant$=k_d$ (driver alone)$+k_B$ (box)
l	effective length of the voice coil wire
m_t	total moving mass, including the air load mass
ω_0	resonance frequency
$\omega_t=1.4c/a$	transient frequency
P	sound pressure
R_e	electrical resistance of the voice coil
R_m	mechanical damping
R_d	electrodynamic damping$=(Bl)^2/R_e$
R_r	real part of $Z_{\text{rad}}=\Re\{Z_{\text{in}}\}$
R_t	total damping$=R_r+R_m+R_d$
ρ	density of the air
s	Laplace variable
S	surface of the cone with radius a
V	velocity of the voice coil
Z_{rad}	mechanical radiation impedance$=R_r+iX_r$

$$H_p(\omega_0)=\frac{P(\omega_0)}{E(\omega_0)}=\frac{i\rho a^2 Bl\omega_0}{2rR_eR_t}. \tag{4}$$

Equation (4) shows that the sensitivity at the resonance frequency depends on the mass m_t via ω_0. Section IV will elaborate on this, and it is shown that at the resonance frequency it is beneficial to have a low-Bl value.

The electrical input impedance can be written (Aarts, 2005) as

$$Z_{\text{in}}(\omega)=R_e\left[1+\frac{Q_{mr}/Q_e}{1+iQ_{mr}\nu}\right]. \tag{5}$$

From this it appears that—via Q_e—Bl plays an important role in the electrical impedance, which is most pronounced at the resonance frequency. By neglecting Z_{rad}, which is very small at low frequencies—in particular at ω_0—the electrical input impedance at ω_0 can be approximated as

$$Z_{\text{in}}(\omega_0)\approx R_e+(Bl)^2/R_m. \tag{6}$$

In order to calculate the power efficiency of loudspeakers, it is required to calculate the electrical power delivered to the driver as well as the acoustical power radiated by the loudspeaker. The latter depends on the radiation impedance of the driver. Below, expressions for these three quantities are derived.

Assuming a sinusoidal driving signal, the time-averaged electrical power P_e delivered to the driver can be written as

$$P_e=0.5|I_c|^2\Re\{Z_{\text{in}}\}=0.5|I_c|^2R_e\left[1+\frac{Q_{mr}/Q_e}{1+Q_{mr}^2\nu^2}\right], \tag{7}$$

where $\Re\{Z_{\text{in}}\}$ is the real (resistive) part of the input impedance Z_{in}. The radiation impedance Z_{rad} of a plane circular rigid piston with a radius a in an infinite baffle can be derived as (Morse and Ingard, 1968, p. 384)

$$Z_{\text{rad}}=\pi a^2\rho c[1-2J_1(2ka)/(2ka)+i2\mathbf{H}_1(2ka)/(2ka)], \tag{8}$$

where $\mathbf{H}_1$ is a Struve function and J_1 is a Bessel function (Abramowitz and Stegun, 1972, Secs. 12.1.7 and 9, respectively), and k is the wave number ω/c. An accurate, full-range approximation of $\mathbf{H}_1$ is given in Aarts and Janssen (2003) as

$$\mathbf{H}_1(z)\approx\frac{2}{\pi}-J_0(z)+\left(\frac{16}{\pi}-5\right)\frac{\sin z}{z}+\left(12-\frac{36}{\pi}\right)\frac{1-\cos z}{z^2}. \tag{9}$$

For low frequencies ($\omega\ll\omega_t=1.4c/a$) the damping influence of Z_{rad} can either be neglected, or the following approximation (Aarts, 2005) can be used:

$$R_r=\Re\{Z_{\text{rad}}\}\approx\pi a^2\rho c(ka)^2/2. \tag{10}$$

Assuming $c=343$ m/s, $\rho=1.21$ kg/m^3, Eq. (10) yields

$$R_r\approx(0.15Sf)^2, \tag{11}$$

where $f=\omega/2\pi$.

The time-averaged acoustically radiated power can be calculated as

$$P_a=0.5|V|^2\Re\{Z_{\text{rad}}\}, \tag{12}$$

which can be written (Aarts, 2005) as

$$P_a=\frac{0.5(Bl/(R_m+R_r))^2I_c^2R_r}{1+Q_{mr}^2\nu^2}. \tag{13}$$

Using Eqs. (7) and (13), the power efficiency can now be calculated as

$$\eta(\nu)=P_a/P_e=[Q_eQ_r(\nu^2+1/Q_{mr}^2)+Q_r/Q_{mr}]^{-1}. \tag{14}$$

This function depends on all loudspeaker parameters and the frequency. In classical loudspeaker design theory the parameters are chosen such that the sensitivity function $H_p(\omega)$ given by Eq. (3) has a flat characteristic for $\omega>\omega_0$, which implies that $Q_t\approx1/\sqrt{2}$. This gives little freedom in the design parameters. Furthermore, one wants a reasonable efficiency. Recently, Vanderkooy *et al.* (2003) investigated the use of high-Bl drivers. The aim of that study was to obtain efficient loudspeakers; however, they have a poor sensitivity at low frequencies. In the following section the use of low-Bl drivers is discussed; those drivers appeared to be highly sensitive and exhibit a good efficiency, but only around the resonance frequency.

III. SPECIAL DRIVERS FOR LOW FREQUENCIES

Two options are described whereby modifying a conventional loudspeaker driver can lead to enhanced bass performance. This is achieved by modifying the force factor of the driver, in particular by employing either a very strong or very weak magnet compared to what is commonly used in typical drivers. Both these approaches also require some preprocessing of the signal before it is applied to the modified loud-

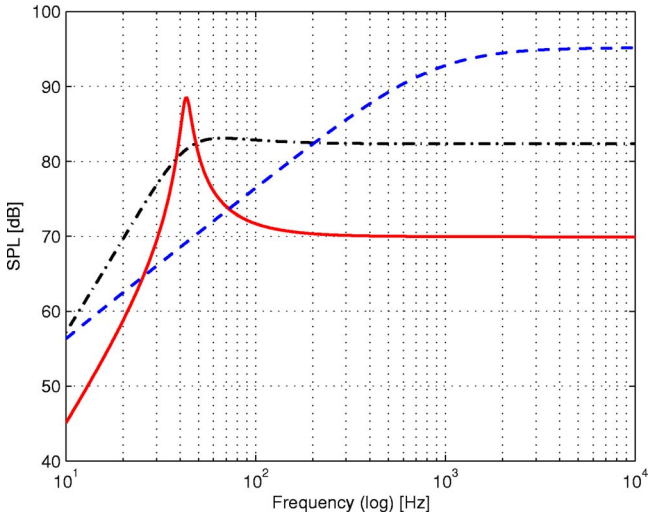


FIG. 1. (Color online) Sound-pressure level (SPL) for the driver MM3c with three Bl values: low $Bl=1.2$ (solid), medium $Bl=5$ (dash-dot), and high $Bl=22$ N/A (dash), while all other parameters are kept the same as given in Table II, all with 1-W input power and at 1 m distance. At the resonance frequency, the highest SPL is obtained by the low-Bl driver, while the high-Bl driver has at low frequencies—in particular at the resonance frequency—a poor response.

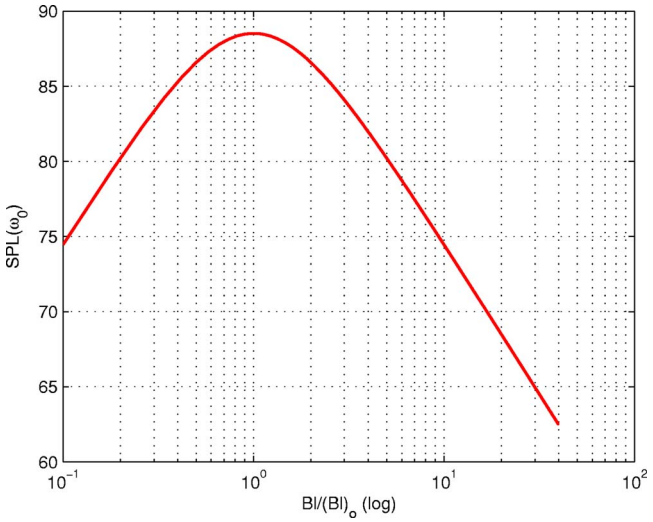


FIG. 2. (Color online) The SPL at the resonance frequency versus the normalized force factor $Bl/(Bl)_o$ for the driver MM3c, where $(Bl)_o$ is the optimal force factor given in Eq. (29), in this present case $(Bl)_o=1.19$. The other parameters are given in Table II.

speaker. In the remaining section the influence of the force factor on the performance of the loudspeaker is reviewed.

Direct-radiator loudspeakers typically have a very low efficiency, because the acoustic load on the diaphragm or cone is relatively low compared to the mechanical load. In addition, the driving mechanism of a voice coil is quite inefficient in converting electrical energy into mechanical motion. The force factor Bl is deliberately kept at an intermediate level so that the typical response is sufficiently flat to use the device without significant equalization. It was already shown in Sec. II that the force factor Bl plays an important role in loudspeaker design. It determines among others the frequency response and its related transient response, the electrical input impedance, and the weight of a loudspeaker; the following will discuss these various characteristics.

To show the influence on the frequency response, the sound-pressure level (SPL) of a driver with three Bl values (low, medium, and high) is plotted in Fig. 1, while all other parameters are kept the same.

It is seen that the curves change drastically for varying Bl. The most prominent difference is the shape, but also apparent is the difference in level at high frequencies. While the low-Bl driver has the highest response at the resonance frequency, it has a poor response beyond resonance, so in use this loudspeaker requires special treatment, as discussed in Sec. IV. The high-Bl driver has a good response at higher frequencies, but a poor response at lower frequencies, which requires special equalization. In between, there is the well-known curve for a medium-Bl driver. The influence of Bl on the sensitivity at the resonance frequency is further clarified by plotting the SPL at the resonance frequency versus the normalized Bl, as is shown in Fig. 2.

It appears that at the resonance frequency there is an optimal value for the voltage sensitivity at $Bl/Bl_o=1$, where Bl_o is the optimal-Bl value discussed in Sec. IV. The underlying reason for the importance of Bl is that, besides determining the driving force, it also provides (electrodynamic) damping to the system. The total damping R_t is equal to the sum of the real part of the radiation impedance R_r, the mechanical damping R_m, and the electrodynamic damping $R_d=(Bl)^2/R_e$, where the electrodynamic damping dominates for medium- and high-Bl loudspeakers, and is most prominent around the resonance frequency. The variables in this electrodynamic damping term cannot be selected independently. This can be seen as follows. The voice coil resistance can be written as

$$R_e=\frac{l\rho_e}{A_e}, \tag{15}$$

where ρ_e and A_e are the electric conductivity and area of the voice coil wire, respectively. The volume occupied by the voice coil is equal to

$$V_e=A_e l. \tag{16}$$

Combining these two equations yields the electrodynamic damping

$$R_d=\frac{(Bl)^2}{R_e}=\frac{B^2V_e}{\rho_e}, \tag{17}$$

which shows that the volume occupied by the voice coil, and the material used for the magnet and voice coil wire, determines the electrodynamic damping, and not the length l of the voice coil's wire.

The power efficiency given in Eq. (14) can be written as

$$\eta(\omega)=\frac{(Bl)^2R_r}{R_e\{(R_m+R_r)^2+(R_m+R_r)(Bl)^2/R_e+(m_t\omega_0\nu)^2\}}. \tag{18}$$

If $(m_t\omega_0\nu)^2\gg[(R_m+R_r)^2+(R_m+R_r)(Bl)^2/R_e]$, and R_r is approximated by Eq. (10), then Eq. (18) can be written as

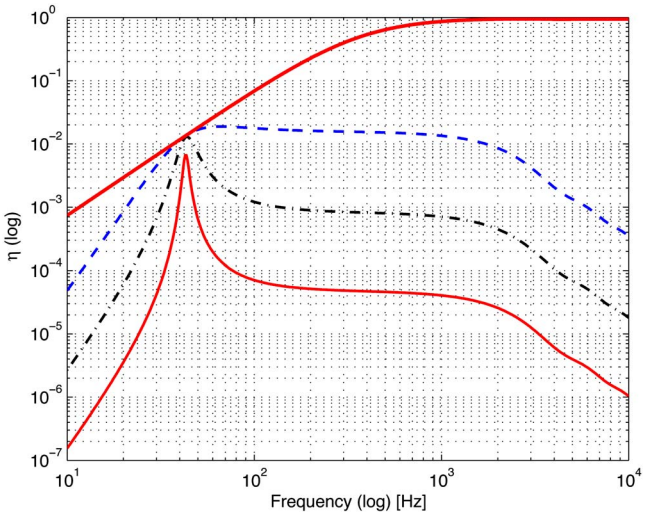


FIG. 3. (Color online) The power efficiency η for the driver MM3c with four Bl values: low $Bl=1.2$ (solid), medium $Bl=5$ (dash-dot), high $Bl=22$ N/A (dash), and $\lim_{Bl\to\infty}$ (thick solid), while all other parameters are kept the same as given in Table II. Note that the efficiency is strongly dependent on Bl at all frequencies except at resonance, where the efficiency is affected only modestly by Bl.

$$\eta(\omega_0 \ll \omega \ll \omega_t) \approx \frac{(Bl)^2 S^2 \rho}{2\pi c R_e m_t^2}. \tag{19}$$

This is a well-known result in the literature (Beranek, 1954) and clearly shows the influence of Bl, however, is valid in a limited frequency range only.

Using Eq. (18), the power efficiency η is plotted in Fig. 3, which clearly shows the dependency on frequency.

Figure 3 shows the efficiency function η as function of the frequency for various values of Bl, while all other parameters are kept the same. It appears that the curves change drastically for varying Bl, but only very modestly around the resonance frequency. This can further be clarified by using Eq. (18) and calculating the limit

$$\lim_{Bl\to\infty} \eta(\omega) = \frac{R_r(\omega)}{R_m + R_r}. \tag{20}$$

The curve for $\eta(\omega)$ for this infinite-Bl value is the thick-solid curve in Fig. 3. Assuming that $R_r \ll R_m$ and $\omega=\omega_0$, and using Eqs. (10) and (20), this yields at the resonance frequency

$$\lim_{Bl\to\infty} \eta(\omega=\omega_0) \approx \frac{R_r(\omega_0)}{R_m} \approx \frac{\rho(S\omega_0)^2}{2\pi c R_m}. \tag{21}$$

Equation (21) shows the approximate value of the power efficiency at the resonance frequency for infinite Bl. It appears that the four curves of Fig. 3 are almost coincident at the point given by Eq. (21), even for the low-Bl curve. This is further elucidated in Fig. 4. This graph shows the power efficiency at the resonance frequency versus Bl/Bl_o, where Bl_o is the optimal-Bl value discussed in Sec. IV.

Figure 4 shows an s curve, where the part for very low-Bl values exhibits a very poor efficiency. There, the Lorentz force acting on the driver's voice coil is small with respect to the damping. Then, a rather steep part of the curve follows,

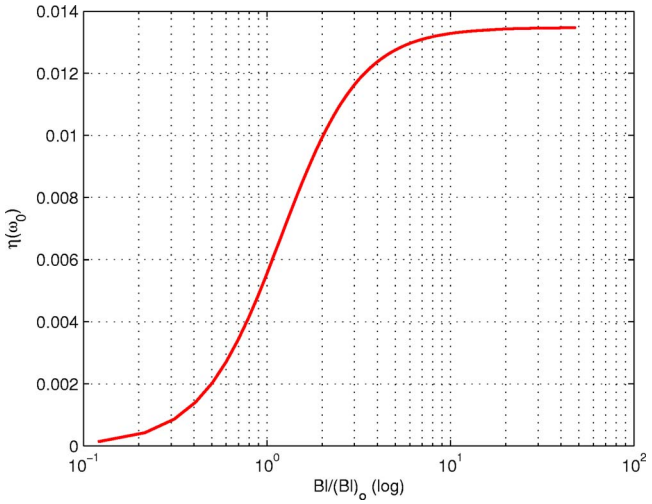


FIG. 4. (Color online) The power efficiency $\eta(\omega=\omega_0)$ versus the normalized force factor $Bl/(Bl)_o$ for the driver MM3c, where $(Bl)_o$ is the optimal force factor given in Eq. (29), in this present case $(Bl)_o=1.19$. The other parameters are given in Table II.

and finally a plateau exists, which is given by Eq. (20). The importance of Bl is further elucidated in the following section.

IV. LOW-FORCE-FACTOR DRIVERS

As explained before, normally low-frequency sound reproduction with small transducers is quite inefficient. Two measures are proposed to increase the efficiency. First, a special transducer is used with a low-Bl value, attaining a high efficiency and the highest possible sensitivity at that particular frequency. Second, nonlinear processing essentially compresses the bandwidth of a 20- to 120-Hz bass signal down to a much narrower span. This span is centered at the resonance of the low-Bl driver where its efficiency is maximum. These drivers are only useful for subwoofers. In the following an optimal force factor is derived to obtain such a result.

The proposed solution, to obtain a high sound output from a compact loudspeaker arrangement with a good efficiency, consists of two steps. First, the requirement that the frequency response must be flat is relaxed. By making the magnet considerably smaller and lighter (see Fig. 5, right

FIG. 5. (Color online) Picture of the prototype driver (MM3c) with a 10 Euro cents coin. At the position where a normal loudspeaker has its heavy and expensive magnet, the prototype driver has an almost empty cavity; only a small moving magnet is necessary, which is shown in the right corner.

TABLE II. The lumped parameters of the new, and experimental driver with the (optimal) low-*Bl* MM3c; see Fig. 5 for its compact magnet system. See Table I for the abbreviations and the meaning of the variables.

Type	MM3c
R_e Ω	6.4
Bl N/A	1.2
k_d N/m	1022
m_t g	14.0
R_m Ns/m	0.22
S cm^2	86
f_0 Hz	43
Q_m	17.2
Q_e	16.8

side) a large peak in the SPL curve [see Fig. 1 (solid curve)] will appear.

Because the magnet can be considerably smaller than usual, the loudspeaker can be of the moving magnet type with a stationary coil (see Fig. 5) instead of vice versa. At the resonance frequency the voltage sensitivity can be a factor of 10 higher than that of a normal loudspeaker. In this case an SPL of almost 90 dB at 1-W input power at 1-m distance is achieved at the resonance frequency, even when using a small cabinet (<1 1). Because it is operating in resonance mode only, the moving mass can be enlarged—which might be necessary owing to the small cabinet—to keep the resonance frequency sufficiently low. This is done without degrading the efficiency of the system because at the resonance frequency $\nu=0$ and the product $m_t\omega_0\nu$ in Eq. (18) becomes equal to zero. See Table II.

Due to the high and narrow peak in the frequency response, the normal operating range of the driver decreases considerably. This makes the driver unsuitable for normal use. To overcome this, a second measure is applied. Nonlinear processing essentially compresses the bandwidth of a 20- to 120-Hz 2.5-octave bass signal down to a much narrower span—which is centered at the resonance of the low-*Bl* driver—where its efficiency is maximum. This can be done with a setup as depicted in Fig. 6 and will be discussed below.

Without loss of generality, it is assumed that the low-frequency part of the music can be modeled as a sinusoid with frequency ω_c which is modulated by a slowly varying signal $m(t)\geqslant 0$. This yields

$$y(t)=[c_m+m(t)]\sin(w_c t), \tag{22}$$

where c_m is a constant, or more precisely

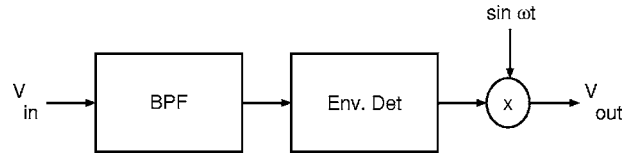


FIG. 6. Frequency mapping scheme. The box labeled "BPF" is a bandpass filter, and "Env. Det." is an envelope detector. The latter can be a simple rectifier followed by a low-pass filter. The signal V_{out} is fed (via a power amplifier) to the driver.

$$h=\frac{\text{peak value of } m(t)}{c_m} \tag{23}$$

is the modulation index. This model is realistic since $y(t)$ is a bandlimited signal, say between 20 to 120 Hz. The frequency of ω_c can be variable and will lead to a certain pitch. Taking the Fourier transform of Eq. (22), the magnitude of the spectrum can be written as

$$|Y(\omega)|=\pi c_m\delta(\omega-\omega_c)+\tfrac{1}{2}M(\omega-\omega_c)+\pi c_m\delta(\omega+\omega_c)+\tfrac{1}{2}M(\omega+\omega_c), \tag{24}$$

where δ is a unit impulse and the capital function $M(\omega)$ indicates the Fourier transform of $m(t)$. Equation (24) shows the well-known amplitude-modulated (AM) spectrum, as known from AM radio broadcasting. In contrast to normal AM radio, in the present case $c_m=0$, this is to make the amplitude of $y(t)$ proportional to the amplitude of $m(t)$. If the processing depicted in Fig. 6 is applied to the signal $y(t)$, the signal $m(t)$ is recovered by an envelope detector and is used to modulate a sinusoid, but now with frequency ω_0, where ω_0 is fixed and equal to the resonance frequency of the transducer. This yields

$$v_{\text{out}}(t)=m(t)\sin(w_0 t), \tag{25}$$

with the corresponding spectrum

$$|V(\omega)|=\tfrac{1}{2}[M(\omega-\omega_0)+M(\omega+\omega_0)]. \tag{26}$$

The result is that the coarse structure $m(t)$ (the envelope) of the music signal after the compression or "mapping" is the same as before the mapping; an example is shown in Fig. 7. Only the fine structure has been changed to a sinusoid of the same frequency as the driver's resonance frequency.

The upper panel shows the waveform of a rock-music excerpt; the thin curve depicts its envelope, $m(t)$. The middle and lower panels show the spectrograms of the input and output signals, respectively, clearly showing that the frequency bandwidth of the signal around 60 Hz decreases after the mapping, yet the temporal modulations remain the same.

Using Eq. (1) and neglecting Z_{rad}, the voltage sensitivity at the resonance frequency can be written as

$$H_p(\omega_0)=\frac{P(\omega_0)}{E(\omega_0)}=\frac{i\omega_0 SBl\rho}{2\pi r R_e(R_m+(Bl)^2/R_e)}. \tag{27}$$

Equation (27) is maximized by adjusting the force factor *Bl* by differentiating $H_p(\omega=\omega_0)$ with respect to *Bl* and setting $\partial H_p/\partial(Bl)=0$, resulting in

$$\frac{(Bl)^2}{R_e}=R_m. \tag{28}$$

It appears that the maximum voltage sensitivity is reached when the electrodynamic damping term $(Bl)^2/R_e$ is equal to the mechanical damping term R_m; in this case the optimal force factor is defined as

$$(Bl)_o=\sqrt{R_e R_m}. \tag{29}$$

The consequences of this optimality criterion are discussed below. One obvious observation is that the SPL response becomes, as can be seen in Fig. 1 (solid curve), very peaky.

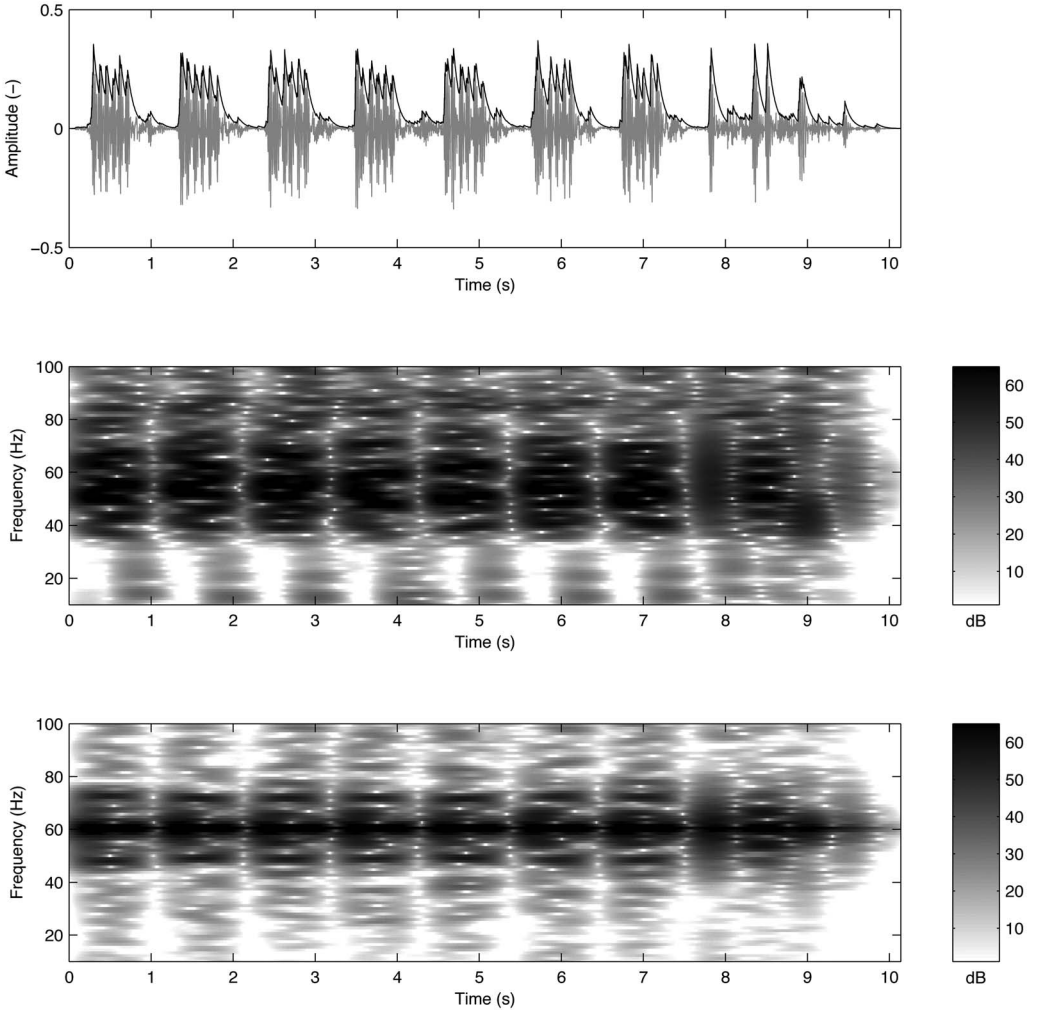


FIG. 7. The signals before and after the frequency-mapping processing of Fig. 6. The upper panel shows the time signal at V_{in}, and the thin curve the output of the envelope detector. The middle and lower panels show the spectrogram of the input and output signals, respectively.

The height of the peak is calculated by substituting Eq. (28) into Eq. (27), which yields the optimal voltage sensitivity

$$H_o(\omega=\omega_0)=\frac{i\omega_0\rho S}{4\pi r(Bl)_o}. \tag{30}$$

The specific relationship between $(Bl)_o$ and both R_m and R_e [Eq. (29)] causes H_o to be inversely proportional to $(Bl)_o$ (which may seem counterintuitive), and thus also inversely proportional to $\sqrt{R_m}$. For this particular value of Bl the Lorentz force is large enough to get a sufficiently strong driver with good efficiency, but the electromagnetic damping is sufficiently low to reach the optimal sensitivity.

The power efficiency at the resonance frequency under the optimality condition, obtained by substitution of Eq. (28) into Eq. (18), yields

$$\eta_o(\omega=\omega_0)=\frac{R_mR_r}{(R_m+R_r)^2+(R_m+R_r)R_m}. \tag{31}$$

This can be approximated for $R_r \ll R_m$ as

$$\eta_o(\omega=\omega_0)\approx\frac{R_r}{2R_m}, \tag{32}$$

which clearly shows that for a high power efficiency at the resonance frequency, the cone area must be large, because R_r—according to Eq. (8), or more explicitly Eq. (11)—is proportional to the squared cone area, and that the mechanical damping must be as small as possible. The damping must be not too small, however, because the transient response depends on the damping as well, as is discussed in Larsen and Aarts (2004). Comparing Eq. (32) with Eq. (21) shows that the optimally sensitive driver is only 3 dB less efficient than the infinite Bl one; however, this is only at the resonance frequency, but this is the working frequency of the new driver. This can also be seen in Fig. 3 where η_0 (solid curve) is close to the infinite Bl curve, but only at the resonance frequency.

V. DISCUSSION

Sound reproduction at low frequencies with small drivers in small cabinets is not efficient. Small drivers have a low radiation impedance with respect to the total damping [see Eq. (18)]. Small cabinets have a stiff air spring which needs

a high moving mass to obtain the desired low resonance frequency. This will be reiterated more quantitatively below. For a given volume of the enclosure V_0, the corresponding k_B of the "air spring" can be calculated as

$$k_B = \frac{\rho(cS)^2}{V_0}. \tag{33}$$

Mounting a loudspeaker in a cabinet will increase the total spring constant k_t by an amount given by Eq. (33), and subsequently increase the resonance frequency of the system. To compensate for this bass loss, the moving mass has to be increased; thus, $\sqrt{k_t m_t}$ is increased, which raises Q_e [see Eq. (2)]. Then—to obtain a flat frequency characteristic—Bl must be increased to preserve the original value of Q_e. This is at the cost of a more expensive magnet and a loss in efficiency. This is the designer's dilemma: high efficiency or small enclosure? To meet the demand for a certain cutoff frequency, the enclosure volume must be larger. Alternatively, the efficiency for a given volume will be less than for a system with a higher cutoff frequency.

This dilemma is (partially) solved by using the low-Bl concept as discussed in Sec. IV, however, at the expense of a decreased sound quality and the need for some additional electronics to accomplish the frequency mapping. While the new driver is not a hi-fi one, many informal listening tests and demonstrations[1] confirmed that the decrease of sound quality appears to be modest, apparently because the auditory system is less sensitive at low frequencies. Also, the other parts of the audio spectrum have a distracting influence on this mapping effect, which has been confirmed during formal listening tests (Le Goff *et al.*, 2004), where the detectability of mistuned fundamental frequencies was determined for a variety of realistic complex signals. Finally, the part of the spectrum which is affected is only between, say, 20 and 120 Hz, so the higher harmonics of these low notes are mostly out of this band and are thus not affected. They will contribute in their normal unprocessed fashion to the missing fundamental effect. All these factors support the notion that detuning becomes difficult to detect once the target complex is embedded in a spectrally and temporally *rich* sound context, as it is typical for applications in modern multimedia reproduction devices (Le Goff *et al.*, 2004).

VI. CONCLUSIONS

The force factor Bl plays a very important role in loudspeaker design. It determines the efficiency, the sensitivity, the impedance, the SPL response, the weight, and the cost. It appears to be not possible to obtain both a high efficiency as well as a high sensitivity in a wide frequency range. At the loudspeaker's resonance frequency, however, it appears to be possible to meet this criterion. The voltage sensitivity is optimal when the electrical damping force is equal to the mechanical one, while it is only 3 dB less efficient than an infinite force-factor loudspeaker. A new low-Bl driver has been developed which together with some additional electronics, yields a low-cost, lightweight, compact, physically flat, optimally sensitive, and very-high-efficiency loudspeaker system for low-frequency sound reproduction.

ACKNOWLEDGMENTS

I would like to thank Joris Nieuwendijk (Philips Applied Technologies) and Okke Ouweltjes (Philips Research), who gave valuable assistance to the low-Bl project.

[1]Demonstrations and MATLAB scripts are on http://www.dse.nl/~rmaarts

Aarts, R. (**2005**). "High-efficiency low-*Bl* loudspeakers," J. Audio Eng. Soc. **53**, 579–592.

Aarts, R., and Janssen, A. (**2003**). "Approximation of the Struve function H1 occurring in impedance calculations," J. Acoust. Soc. Am. **113**, 2635–2637.

Abramowitz, M., and Stegun, I. (**1972**). *Handbook of Mathematical Functions* (Dover, New York).

Beranek, L. (**1954**). *Acoustics* (McGraw-Hill, New York) (Reprinted by ASA 1986).

Larsen, E., and Aarts, R. (**2004**). *Audio Bandwidth Extension. Application of Psychoacoustics, Signal Processing and Loudspeaker Design* (Wiley, New York).

Le Goff, N., Aarts, R., and Kohlrausch, A. (**2004**). "Thresholds for hearing mistuning of the fundamental component in a complex sound," in Proceedings of the 18th International Congress on Acoustics (ICA 2004), Paper Mo.P3.21, p. I-865 (Kyoto, Japan).

Morse, P., and Ingard, K. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).

Small, R. (**1972**). "Closed-box loudspeaker systems. I. Analysis," J. Audio Eng. Soc. **20**, 798–808.

Thiele, A. (**1971**). "Loudspeakers in vented boxes. I," J. Audio Eng. Soc. **19**, 382–392.

Vanderkooy, J., Boers, P., and Aarts, R. (**2003**). "Direct-radiator loudspeaker systems with high *Bl*," J. Audio Eng. Soc. **51**, 625–634.

A method for structural vibration characterization in the midfrequency region[a)]

W. Steve Shepard, Jr.[b)] and Yi Liu
Department of Mechanical Engineering, The University of Alabama, 290 Hardaway Hall, Box 870276, Tuscaloosa, Alabama 35487

(Received 29 September 2004; revised 21 November 2005; accepted 23 November 2005)

The ability to characterize elastic structures at higher frequencies in a meaningful and convenient manner has been a topic of much interest. Many analytical or experimental approaches incorporate the use of lumped quantities to represent critical system characteristics at discrete locations. When the excitation frequency increases, the structural wavelength becomes comparable to the dimensions of that contact region and the point-quantity assumption is no longer valid. To provide a similar means for describing structures at higher frequencies, the work described here reformulates the traditional four-pole method in terms of quantities defined over planes. Spatial variations of the important response characteristics across the connecting regions can be considered. As for the four-pole method, the approach has the advantage of providing building blocks for simple structures in order to represent more complex structures. Three application examples are presented. One example involves modeling a simple vibration isolator. In another example, the impact of having spatial functions that extend over only a portion of the structure is studied. Finally, a two-dimensional structure is examined using a finite element approach. In using the approach, additional insight can be gained that cannot be easily found using typical finite element method analysis results. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151827]

PACS number(s): 43.40.At, 43.40.Cw, 43.40.Tm [JGM] Pages: 897–908

I. INTRODUCTION

The ability to characterize the dynamic response of elastic structures in a convenient and meaningful manner has been a research topic of much interest. Knowledge of the frequency dependent vibration properties of structures is necessary for predicting vibration transmission energy. Furthermore, vibration characterization of coupled structures is vital for acoustic radiation studies such as for naval surface ships and submarines.

A well-developed method for describing the dynamic characteristics of structures, commonly referred to as the four-pole method,[1,2] involves representing the input and output forces and velocities in terms of point quantities. Figure 1(b) illustrates the lumped parameter model for the vibration isolator shown in Fig. 1(a). With the convenient four-pole model and the assumption of lumped physical quantities, one can accurately and readily quantify the dynamic characteristics of an elastic structure, such as for a vibration isolator at low frequencies.[3,4] As the frequency increases, though, the four-pole method is no longer applicable for characterizing the structure since spatial variations of the response characteristics must be considered as frequency increases. Cremer studied the impact of spatial variations in the excitation quantities for an infinite plate.[5] He found that the assumption of point quantities is valid so long as the dimensions of the contact region are less than approximately one-tenth of the structural wavelength for a vibrating plate. At higher frequencies, where the dimensions of the region of contact between two structures become comparable to the wavelength, the spatial characteristics of the dynamic properties must be considered.[6] Such variations in the response are illustrated by the spatially distributed forces and velocities shown in Fig. 1(c). A convenient method for characterizing structures in the "midfrequency" region, where these types of spatial variations occur, is needed. Here "midfrequency" refers to the frequencies where the dimension of the contact region is comparable to the vibration wavelength. It is worth noting that for extreme high frequencies, statistical energy analysis[7] is a commonly used approach. Consequently, the high frequency region is not addressed in this present work. Of course, discretization methods, such as the finite element method, can be used to predict the response in the midfrequency region. These standard solution methods, though, are not so convenient if one is interested in gaging the relative importance of the various dynamic characteristics for the structure. Furthermore, the solutions generally obtained from discretization methods do not provide physical insight regarding the important characteristics of a spatially distributed response.

In this work, a method referred to as the four-plane method will be developed by extending the formulation of the widely accepted four-pole method. This method can be used to describe spatial variations over the contact region for the response quantities of interest. By building on the four-pole parameter method, many of the same useful features for that method are retained. For example, the system character-

[a)]Portions of this work were presented in "A method for characterizing elastic structures in the midfrequency region," Proceedings of 2004 ASME International Mechanical Engineering Congress, Anaheim, California, November 2004.

[b)]Author to whom all correspondence should be addressed. Electronic mail: sshepard@eng.ua.edu

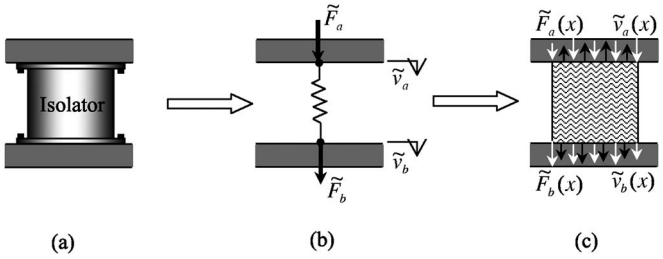


FIG. 1. Structure containing vibration isolator, (a) Actual isolator, (b) low frequency lumped-parameter model, and (c) midfrequency spatial model.

istics of connected structures can be formulated by assembling the representations of the basic elements within that system. The theory for this approach is presented next. Then, some methods needed to formulate the required matrices that are used in the approach are discussed. To demonstrate the applicability of the technique developed here as well as note some of the benefits of the approach, the four-plane method will be applied to a simplified vibration isolator model, a simply supported beam, and a clamped plate. In these examples, various aspects of the results will be noted.

II. THEORY

In order to derive a characterization method for response quantities defined over a finite sized region, first consider the spatially varying forces acting on the structure, as illustrated in Fig. 2. Although not explicitly shown in the figure, the velocities can also vary across the same region of the structure. This structure could be a vibration isolation mount, for example. First consider a description of the input force distribution $\tilde{F}_a$ defined over a planar area. Note that a more general nonplanar contact region may be used. For the sake of simplicity, however, a planar contact region will be assumed here. To describe the force, a windowed truncated series expansion can be used

$$\tilde{F}_a(x,y) \approx \sum_{k=L}^{N} \tilde{F}_{ak} f_k(x,y), \tag{1}$$

where $\tilde{F}_{ak}$ are the weighting coefficients of the spatial orthonormal basis functions f_k that vary with position (x,y). For example, the set of functions f_k may include Fourier functions as well as uniform and linearly varying functions that span the contact region. On the other hand, the function set could also be based on modal functions that span the entire structure. Some examples of these functions will be illustrated in the applications. More typical basis functions are used in the case studies presented later. Although one could use other series to expand function $\tilde{F}_a$, a series of orthonormal functions is used in the examples provided below in order to circumvent mathematical complications. The "≈" in Eq. (1) denotes that the input force is approximated using a finite series expansion. As N increases and L decreases, the level of detail that can be described by the spatial representation increases. It is assumed that the function with the lowest index, $k=1$, is the most simplistic function. For example, f_1 might be the uniform function. By using this notation with $k=L,\ldots,N$, however, it is not necessary to always include such simplistic terms in the expansion. This approach is analogous to using a windowed set of modal functions to represent the response of a structure. Various research has shown that moment effects can be important at higher frequencies.[8] Momentlike excitations can be described in this method by using the appropriate functions in Eq. (1). An example of a momentlike excitation will be presented later in this work. A similar spatial representation with F_{bi} coefficients can be used to describe the output force F_b illustrated in Fig. 2. Although the input and output velocity distributions, $\nu_a(x,y)$ and $\nu_b(x,y)$, can be described using this same representation given in Eq. (1), except with $\tilde{V}_{am}$ and $\tilde{V}_{bj}$ coefficients, respectively, there is no restriction that the basis functions for these expansions be the same. Here, the output force is described using a set of orthonormal functions g_i. The velocity functions will be described later. It is worth noting that a different number of terms may be used in each of these series and any convenient coordinate system may be used.

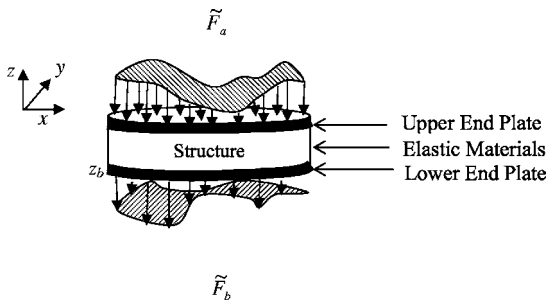


FIG. 2. Structure excited under distributed forces.

To describe the transfer characteristics of the structure, the input and output quantities must be related. Suppose first that the output velocity, $\tilde{v}_b$, is zero. Then, the spatially varying input force can be related to the spatially varying output force by

$$\tilde{\mathbf{F}}_a = \tilde{\mathbf{M}}^{FF}\tilde{\mathbf{F}}_b, \tag{2}$$

where $\tilde{\mathbf{F}}_a$ and $\tilde{\mathbf{F}}_b$ are the complex force weighting coefficient column vectors

$$\tilde{\mathbf{F}}_a = \{\tilde{F}_{aL}\cdots\tilde{F}_{aN}\}^T \tag{3}$$

and

$$\tilde{\mathbf{F}}_b = \{\tilde{F}_{bL}\cdots\tilde{F}_{bN}\}^T. \tag{4}$$

Each $\tilde{F}_{ak}$ and $\tilde{F}_{bi}$ $(i=L,\ldots,N)$ entry for these vectors are the complex weighting coefficients for the f_k and g_i terms in the series expansion, respectively. The superscript T represents the matrix transpose. In Eq. (2), the matrix $\tilde{\mathbf{M}}^{FF}$ describes the coupling between the two spatially varying force quantities and is therefore denoted by the FF superscript. For example, the element located in the first row and fourth column of this matrix, $\tilde{M}^{FF}_{14}$, describes the contribution of a unit amplitude of $\tilde{F}_{b4}$ to the coefficient $\tilde{F}_{a1}$. In general, $\tilde{F}_{a1}$ also depends on contributions from the other force terms, $\tilde{F}_{bi}$, as well as the output velocity. As a result, the full description of the structure behavior in terms of the expansion coefficients

and the relationships between those expansions, is given by

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \tilde{\mathbf{M}}^{FF} & \tilde{\mathbf{M}}^{FV} \\ \tilde{\mathbf{M}}^{VF} & \tilde{\mathbf{M}}^{VV} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}, \tag{5}$$

where the other $\tilde{\mathbf{M}}$ submatrices relate the appropriate force and velocity terms. The $\tilde{\mathbf{M}}^{VF}$ matrix relates each input velocity term to the corresponding output force term. The full $[\tilde{\mathbf{M}}]$ matrix, which contains all the $\tilde{\mathbf{M}}$ submatrices on the right-hand side of the above equation, fully describes the spatial transmission characteristics of the structure. Note that if the same number of terms is used in the expansions to describe the input and output quantities, then $[\tilde{\mathbf{M}}]$ is a square matrix. Equation (5) is in a form that is similar to the four-pole parameter equation[1]

$$\begin{Bmatrix} \tilde{F}_a \\ \tilde{V}_a \end{Bmatrix} = \begin{bmatrix} \tilde{\alpha}_{11} & \tilde{\alpha}_{12} \\ \tilde{\alpha}_{21} & \tilde{\alpha}_{22} \end{bmatrix} \begin{Bmatrix} \tilde{F}_b \\ \tilde{V}_b \end{Bmatrix}, \tag{6}$$

except the discrete parameters $\tilde{\alpha}$ are replaced here with $\tilde{\mathbf{M}}$ submatrices. $\tilde{F}$ and $\tilde{V}$ are point-based quantities in Eq. (6). The model described by Eq. (5) relates four spatially varying quantities defined over planes. As a result, this approach is referred to as the "four-plane" model. Even though the force and velocity are applied over the same geometric plane, their distributions are generally different. To that end, "four-plane" is used to denote different physical quantities over the input and output planes. The application of this terminology is consistent with that used for the four-pole technique. If one term is used in Eq. (1) and the first function is assumed to be a uniform function, then this model becomes the four-pole parameter model and the submatrices become the regular four-pole $\hat{\alpha}$ parameters.[2] At this point it is worth noting that for most applications, only the coupling between a limited number of input and output basis functions is of interest. Therefore, the right-hand side of Eq. (1) represents the projection of $\tilde{F}_a$ in the space spanned by the basis functions of interest. This also applies to other forces and velocities. Of course, in order to obtain an exact characterization of the structure, an infinite number of terms is needed for each series [e.g., $N=\infty$ in Eq. (1)]. Since this is not practical, a windowed series is used in most cases to accurately approximate the description.

Since the four-plane method described here is an extension of the four-pole method, some of the advantages of using the latter method in characterizing elastic structures extend to this approach. First, the $\tilde{\mathbf{M}}$ matrices characterize only the structure for which they are determined. In other words, their values are not influenced by the preceding or subsequently attached mechanical systems.

Based on the formulation, it is apparent from Eq. (5) that the output force and velocity influence the input force. General formulation techniques, therefore, become inherently difficult to use when generating the required matrices due to the coupled nature of these quantities. Consequently, an alternate formulation, such as a modal method, finite element (FE) method, or experimental approach is needed to determine the corresponding terms of the $\tilde{\mathbf{M}}$ submatrices. Illustrated in the following section is the development of a closed-form four-plane solution for finding the coupling matrices using the modal method. Then, the approach is applied to a vibration isolator modeling example in a later section. From this and other examples, some important properties of the $\tilde{\mathbf{M}}$ submatrices will become evident.

III. $\tilde{\mathbf{M}}$ MATRIX MODAL METHOD FORMULATION

In this section, an analytical form of the $\tilde{\mathbf{M}}$ submatrices for a finite-length elastic structure is obtained using the modal superposition technique. Note the structures considered here are assumed to have no loss for the sake of simplicity. For a linear structure under a distributed force $\tilde{F}_a$, the velocity response $\tilde{\nu}$ can be expressed as

$$\tilde{\nu} = \mathrm{j}\omega \sum_{t=1}^{n} \tilde{q}_t \phi_t, \tag{7}$$

where ϕ_t is the modal shape function. The $\tilde{\nu}$ function defined here extends over the entire structure and not just the contact regions. The generalized coefficient $\tilde{q}_t$ is the modal coefficient given by

$$\tilde{q}_t = \frac{\langle F_a, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}, \tag{8}$$

where m_t is the modal mass, ω_t is the angular natural frequency for the tth mode, and ω is the excitation frequency. Note that this relationship assumes there are no other forces acting on the structure. The inner product operator $\langle F_a, \phi_t \rangle$ is defined as

$$\langle F_a, \varphi_i \rangle = \int_{\Omega} F_a \varphi_i d\Omega, \tag{9}$$

where Ω is the contact region, which can be either the input area "a" or the output area "b." If F_a is expanded in a series form as given by Eq. (1) and one assumes that F_a is a conphase distributed force for the sake of simplicity, Eq. (8) can be transformed to

$$\tilde{q}_t = \frac{\sum_{k=1}^{N} \tilde{F}_{ak} \langle f_k, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}. \tag{10}$$

For the special case where $\tilde{F}_a = \tilde{F}_{ak} f_k$, which corresponds to one term for Eq. (1), the distributed velocity function $\tilde{\nu}_k$ corresponding to this function f_k is

$$\tilde{\nu}_k = \mathrm{j}\omega \tilde{F}_{ak} \sum_{t=1}^{n} \frac{\langle f_k, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)} \phi_t. \tag{11}$$

As noted above, the velocity distribution $\tilde{\nu}_{kb}$ defined over the output contact area b can also be represented using a series expansion

$$\tilde{\nu}_{kb} = \sum_{j=1}^{N} \tilde{V}_{bj} g_j(x,y), \tag{12}$$

where $\tilde{V}_{bj}$ are the complex weighting coefficients and g_j is the orthonormal spatial function defined over region b. The set of functions g_j may include Fourier functions, for example. It is clear that the spatially varying velocity $\tilde{\nu}_{kb}$ is a subset of $\tilde{\nu}_k$ since $\tilde{\nu}_{kb}$ only has nonzero values over region b. As a result

$$\int_b g_j \cdot \tilde{\nu}_k dxdy \Big|_{z=z_b} = \int_b g_j \cdot \tilde{\nu}_{kb} dxdy \Big|_{z=z_b}. \tag{13}$$

Note here the notation x, y, z, and z_b is illustrated in Fig. 2. Substituting Eq. (11) into Eq. (13) and using the orthonormal property of g_j, Eq. (13) can be simplified as

$$\tilde{V}_{bj} = \tilde{F}_{ak} \cdot \mathrm{j}\omega \cdot \sum_{t=1}^{n} \frac{\langle f_k, \phi_t \rangle \langle g_j, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}. \tag{14}$$

This equation gives an analytical relationship between the force coefficient $\tilde{F}_{ak}$ and the velocity coefficient $\tilde{V}_{bj}$. Note if f_k and g_j are selected to be the normalized modal functions, then

$$\sum_{t=1}^{n} \frac{\langle f_k, \phi_t \rangle \langle g_j, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)} = \begin{cases} \dfrac{\langle \phi_k, \phi_k \rangle^2}{m_k(\omega_t^2 - \omega^2)} & k = j \\ 0 & k \neq j \end{cases} \tag{15}$$

due to the orthogonal properties of the functions. From Eqs. (14) and (15), if multiple terms are considered for the force and velocity representations, one can express the relationship between the force and velocity terms in a matrix form as

$$\begin{Bmatrix} \tilde{\mathbf{V}}_a \\ \tilde{\mathbf{V}}_b \end{Bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{C}} & \tilde{\mathbf{D}} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_a \\ -\tilde{\mathbf{F}}_b \end{Bmatrix}, \tag{16}$$

where $\tilde{\mathbf{F}}_a$, $\tilde{\mathbf{F}}_b$, $\tilde{\mathbf{V}}_a$, $\tilde{\mathbf{V}}_b$ are the same as those given in Eq. (5). Note that the minus sign before $\tilde{\mathbf{F}}_b$ is due to the fact the output force is in a direction opposite to that of a typical reaction force. Also note that each entry C_{jk} in the $\tilde{\mathbf{C}}$ matrix can be represented using Eq. (14) as

$$C_{jk} = \mathrm{j}\omega \cdot \sum_{t=1}^{n} \frac{\langle g_j, \phi_t \rangle \cdot \langle f_k, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}, \tag{17}$$

since that term describes the relationship between the $\tilde{\mathbf{F}}_a$ and $\tilde{\mathbf{V}}_b$ coefficients. Using the same approach for the other relationships, each entry of the remaining matrices $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, and $\tilde{\mathbf{D}}$ can be determined from the relationships

$$A_{mk} = \mathrm{j}\omega \cdot \sum_{t=1}^{n} \frac{\langle f_m, \phi_t \rangle \cdot \langle f_k, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}, \tag{18}$$

$$B_{mi} = \mathrm{j}\omega \cdot \sum_{t=1}^{n} \frac{\langle f_m, \phi_t \rangle \cdot \langle g_i, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}, \tag{19}$$

and

$$D_{ji} = \mathrm{j}\omega \cdot \sum_{t=1}^{n} \frac{\langle g_j, \phi_t \rangle \cdot \langle g_i, \phi_t \rangle}{m_t(\omega_t^2 - \omega^2)}. \tag{20}$$

These matrices describe the spatial transmission characteristics of the structure for a linear structure allowing superposition and can be regarded as mobility matrices. Although these matrices are obtained here using a modal method, in practice one can also use a FE method or experimental tools to derive these mobility matrices. This approach is similar to that for the four-pole method. Of course, it may be necessary in such cases to use force reconstruction methods[9] or some other techniques to measure the spatially distributed forces. However, this topic is beyond the scope of this work and will not be discussed here. The example given in the following section will show how to obtain the mobility matrices using FE methods.

Since the ultimate goal of this section is to obtain analytical relationships for the $\tilde{\mathbf{M}}$ submatrices in Eq. (5), Eq. (16) can first be partitioned and rearranged in the form

$$\begin{bmatrix} \tilde{\mathbf{A}} & -\mathbf{I} \\ -\tilde{\mathbf{C}} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} -\tilde{\mathbf{B}} & \mathbf{0} \\ \tilde{\mathbf{D}} & -\mathbf{I} \end{bmatrix} \begin{Bmatrix} -\tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}, \tag{21}$$

where $\mathbf{I}$ denotes the identity matrix. The next step is to premultiply this equation by the inverse of the matrix on the left-hand side to obtain

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}} & -\mathbf{I} \\ -\tilde{\mathbf{C}} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -\tilde{\mathbf{B}} & \mathbf{0} \\ \tilde{\mathbf{D}} & -\mathbf{I} \end{bmatrix} \begin{Bmatrix} -\tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}. \tag{22}$$

The left-most matrix on the right-hand side of Eq. (22) can be simplified as

$$\begin{bmatrix} \tilde{\mathbf{A}} & -\mathbf{I} \\ -\tilde{\mathbf{C}} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{0} & -\tilde{\mathbf{C}}^{-1} \\ -\mathbf{I} & -\tilde{\mathbf{A}}\tilde{\mathbf{C}}^{-1} \end{bmatrix} \tag{23}$$

since the lower right matrix partition is the zero matrix. Using this identity, Eq. (22) takes the final form

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{D}} & -\tilde{\mathbf{C}}^{-1} \\ -\tilde{\mathbf{B}} + \tilde{\mathbf{A}}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{D}} & \tilde{\mathbf{A}}\tilde{\mathbf{C}}^{-1} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}. \tag{24}$$

In comparing Eq. (24) to Eq. (5), the closed-form representation of the $\tilde{\mathbf{M}}$ submatrices in terms of the $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{D}}$ submatrices can be obtained as

$$\tilde{\mathbf{M}}^{FF} = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{D}}, \tag{25}$$

$$\tilde{\mathbf{M}}^{FV} = -\tilde{\mathbf{C}}^{-1}, \tag{26}$$

$$\tilde{\mathbf{M}}^{VF} = -\tilde{\mathbf{B}} + \tilde{\mathbf{A}}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{D}}, \tag{27}$$

and

$$\tilde{\mathbf{M}}^{VV} = \tilde{\mathbf{A}}\tilde{\mathbf{C}}^{-1}. \tag{28}$$

Therefore, by using superposition to compute the velocity response of the structure for the multiple applied forces, the resulting relationships can be reformulated to suit the needs

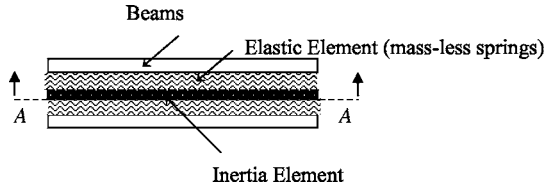


FIG. 3. Simplified vibration isolator model.

of the characterization technique developed here.

In the following section, the four-plane method will first be applied to a simplified vibration isolator model at frequencies higher than those of applicability for the four-pole method. In this case, the contact area is larger than one-tenth of the structural wavelength. The purpose of this example is to demonstrate the potential applications of the four-plane method and to illustrate some useful properties of the full $[\tilde{\mathbf{M}}]$ matrix. In addition, the application of the four-plane method on this vibration isolator model will also demonstrate the potential use of this technique to characterize a vibration isolator at frequencies higher than those typically considered. While standard FE method formulations could be used for the entire modeling process at higher frequencies, the typical form of the discretized response does not provide a foundation against which further analytical insight can be easily gained. The approach presented here has the potential to be used for gaining insight into the relative importance of various response characteristics of the structure even when used in conjunction with using these discretized approaches. It may also be possible to carry this technique with other methods to gain further insight. Such approaches will be the topics of future research. Before proceeding, note that when the contact region between two structures is small compared with the size of the connected structures, but not necessarily small relative to the wavelength, it is more attractive to expand the force and velocity in Eq. (1) using local basis functions rather than modal functions. To that end, the impact of the choice of different basis functions in Eq. (1) on the condition of the full $[\tilde{\mathbf{M}}]$ matrix will also be examined using computational examples.

IV. CASE STUDIES

To demonstrate the applicability of the four-plane method, several examples will be presented. First, the formulation for three basic components will be formulated. These components include a distributed locally reacting spring, a distributed mass, and a free-free beam. Once the matrices for these components are obtained, they will be used to construct the four-plane model for a simplified vibration isolator, as illustrated in Fig. 3. Note the upper mount and below mount plates in the isolator will be represented using the free-free beams. By using this building block approach, the benefits obtained by extending the widely accepted four-pole approach become apparent. Once these basic examples have been presented, slightly more complicated cases are considered so that some of the results of this approach can be examined in more detail.

A. Four-plane matrices for massless distributed spring elements

For massless springlike elements, the input force distribution equals to the output force distribution. Furthermore, a sinusoidal force can be related with velocities by using a distributed stiffness $k(x)$, where only variations in the x direction are considered for this particular problem. The relationship for this locally reacting stiffness is given by

$$\tilde{F}_a(x) = \tilde{F}_b(x), \tag{29}$$

and

$$\tilde{F}_a(x) = \frac{k(x)}{\omega \mathrm{j}}[\tilde{v}_a(x) - \tilde{v}_b(x)], \tag{30}$$

where the complex notation $\mathrm{j} = \sqrt{-1}$ is used. Note that this nonitalic "j" should not be confused with the italic "j" index used above. Rearranging Eqs. (29) and (30) in a four-plane form and assuming that $k(x)$ is a constant value for x, it can be shown that

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{I}\dfrac{\omega \mathrm{j}}{k} & \mathbf{I} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix} = [\mathbf{M}]_{\text{spring}} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix} \tag{31}$$

for the distributed spring element.

B. Four-plane matrices for distributed inertia elements

For a distributed inertia element with an infinite local stiffness, the forces and velocities can be related as

$$\tilde{F}_a(x) = \tilde{F}_b(x) + \rho\omega \mathrm{j} \cdot \tilde{v}_b(x), \tag{32}$$

and

$$\tilde{v}_a(x) = \tilde{v}_b(x), \tag{33}$$

where ρ is the mass density of the element. Note that a locally reactive mass is assumed. Equations (32) and (33) can be rearranged in a four-plane form

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \cdot \rho\omega \mathrm{j} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix} = [\mathbf{M}]_{\text{inertia}} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}. \tag{34}$$

C. Four-plane matrices for a free-free beam

In formulating the matrices for a free-free beam, the input and output basis functions are selected to be the modal shape functions for that beam. According to Eqs. (15), (17)–(20), and (24), the four-plane model for this free-free beam can be given as

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \mathbf{I} & -\tilde{\mathbf{A}}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix} = [\mathbf{M}]_{\text{beam}} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}, \tag{35}$$

where matrix $\tilde{\mathbf{A}}$ is given by

$$\tilde{\mathbf{A}} = \begin{bmatrix} \ddots & & \\ & \dfrac{\langle \phi_t, \phi_t \rangle^2 \omega}{m_t(\omega_t^2 - \omega^2)}\mathrm{j} & \\ & & \ddots \end{bmatrix}. \tag{36}$$

As before, ϕ_t is the tth modal shape function. Note that for this formulation, the input and output quantities extend over the entire upper and lower surface of the beam and thickness modes are neglected. If the contact region extends only over a limited portion of the beam, then the resulting relationships would be different than those given here. Now that the four-plane matrices for each subcomponent have been determined, it is possible to formulate the matrices for a more complicated structure using the results from these components.

D. Four-plane matrices for a simplified vibration isolator

As stated above, one advantage of the traditional four-pole method is its connectivity property.[1] The connectivity property states that a complex system can be regarded as a combination of the respective components using simple matrix algebra. This property is due to the fact that the output of one component is the input of the subsequently connected component. In connecting the poles of the individual components in the lumped-parameter formulation, one merely performs a matrix operation. This property can be extended to the four-plane formulation. As a result, the four-plane formulation for the more complex system, illustrated in Fig. 3, is

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = [\mathbf{M}]_{\text{beam}}[\mathbf{M}]_{\text{spring}}[\mathbf{M}]_{\text{inertia}}[\mathbf{M}]_{\text{spring}}[\mathbf{M}]_{\text{beam}} \begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix}, \tag{37}$$

where $[\mathbf{M}]_{\text{beam}}$, $[\mathbf{M}]_{\text{spring}}$, and $[\mathbf{M}]_{\text{intertia}}$ matrices represent the four-plane matrices of the beam, springlike elements, and inertia elements, respectively. Note that in Eq. (37), the basis functions for each element must be the same to ensure the validity of the matrix multiplicity. Otherwise, a transformation matrix is needed to relate the two matrices constructed using different basis functions. The matrices in Eq. (37) are given in Eqs. (31), (34), and (35). Substituting these into Eq. (37), the four-plane model for the isolator is

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \tilde{\mathbf{M}}^{FF} & \tilde{\mathbf{M}}^{FV} \\ \tilde{\mathbf{M}}^{VF} & \tilde{\mathbf{M}}^{VV} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}, \tag{38}$$

where

$$\tilde{\mathbf{M}}^{FF} = \mathbf{I}\left(1 + \frac{\rho\omega^2}{k}\right) + \tilde{\mathbf{A}}^{-1}\left(\frac{2\omega\mathrm{j}}{k} + \frac{\rho}{k^2}\omega^3\mathrm{j}\right), \tag{39}$$

$$\tilde{\mathbf{M}}^{FV} = -\tilde{\mathbf{A}}^{-2}\left(\frac{2\omega\mathrm{j}}{k} + \frac{\rho}{k^2}\omega^3\mathrm{j}\right) + \mathbf{I}\cdot\rho\omega\mathrm{j} - 2\tilde{\mathbf{A}}^{-1}\left(1 + \frac{\rho\omega^2}{k}\right), \tag{40}$$

$$\tilde{\mathbf{M}}^{VF} = -\mathbf{I}\left(\frac{2\omega\mathrm{j}}{k} + \frac{\rho}{k^2}\omega^3\mathrm{j}\right), \tag{41}$$

and

$$\tilde{\mathbf{M}}^{VV} = \mathbf{I}\left(1 + \frac{\rho\omega^2}{k}\right) + \tilde{\mathbf{A}}^{-1}\left(\frac{2\omega\mathrm{j}}{k} + \frac{\rho}{k^2}\omega^3\mathrm{j}\right). \tag{42}$$

Equations (39) to (42) can be reformulated into the form

$$\tilde{\mathbf{M}}^{FF} = \mathbf{I}\left[\left(1 + \frac{\omega^2}{\omega_s^2}\right) + (\omega^2 - \omega_t^2)\left(\frac{2}{\omega_b^2} + \frac{\omega^2}{\omega_b^2\omega_s^2}\right)\right], \tag{43}$$

$$\tilde{\mathbf{M}}^{FV} = \mathbf{I}\left[\frac{\rho_b(\omega_t^2 - \omega^2)^2}{\omega}\left(\frac{2}{\omega_b^2} + \frac{\omega^2}{\omega_b^2\omega_s^2}\right) + \rho\omega - \frac{2\rho_b(\omega_t^2 - \omega^2)}{\omega}\left(1 + \frac{\omega^2}{\omega_s^2}\right)\right]\mathrm{j}, \tag{44}$$

$$\tilde{\mathbf{M}}^{VF} = -\mathbf{I}\frac{\omega}{\rho_b}\left(\frac{2}{\omega_b^2} + \frac{\omega^2}{\omega_b^2\omega_s^2}\right)\mathrm{j}, \tag{45}$$

and

$$\tilde{\mathbf{M}}^{VV} = \mathbf{I}\left[\left(1 + \frac{\omega^2}{\omega_s^2}\right) + (\omega^2 - \omega_t^2)\left(\frac{2}{\omega_b^2} + \frac{\omega^2}{\omega_b^2\omega_s^2}\right)\right], \tag{46}$$

where ρ_b is the density of beam, $\omega_b = \rho_b/k$, and $\omega_s = \rho/k$. Examination of these $\tilde{\mathbf{M}}$ submatrices reveals that

$$\tilde{\mathbf{M}}^{FF} = \tilde{\mathbf{M}}^{VV}. \tag{47}$$

Furthermore, it can be shown that

$$\begin{vmatrix} \tilde{\mathbf{M}}^{FF} & \tilde{\mathbf{M}}^{FV} \\ \tilde{\mathbf{M}}^{VF} & \tilde{\mathbf{M}}^{VV} \end{vmatrix} = 1. \tag{48}$$

Equations (47) and (48) illustrate that the system is symmetric and satisfies the reciprocity theorem,[2] which follows directly from the conventional four-pole method. It should be noted that in this example, these properties are valid due to the fact that the input and output basis functions are selected to be the full modal functions of one component. For other cases, like one of the examples given in the following section, these properties might not apply.

In order to study the interstructure dynamics, the four-plane equation for half of the system can be considered. For this case, the system is cut along the *A-A* line as shown in Fig. 3. The four-plane formulation for only the part of the system above the dot is

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = [\mathbf{M}]_{\text{beam}}[\mathbf{M}]_{\text{spring}}[\mathbf{M}]_{\text{inertia}} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}. \tag{49}$$

or

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \mathbf{I} + \frac{\rho\omega^2}{k}\mathbf{I} & -\tilde{\mathbf{A}}^{-1} - \tilde{\mathbf{A}}^{-1}\frac{\rho\omega^2}{k} + \mathbf{I}\cdot\rho\omega\mathrm{j} \\ -\mathbf{I}\frac{\omega\mathrm{j}}{\mathrm{k}} & \tilde{\mathbf{A}}^{-1}\frac{\omega\mathrm{j}}{\mathrm{k}} + \mathbf{I} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}. \tag{50}$$

Eq. (50) can be reformulated into

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix} = \begin{bmatrix} \mathbf{I}\left(1 + \frac{\omega^2}{\omega_s^2}\right) & \mathbf{I}\left[\frac{\rho_b(\omega^2 - \omega_t^2)}{\omega}\frac{\omega^2 + \omega_s^2}{\omega_s^2} + \rho\omega\right]\mathrm{j} \\ -\mathbf{I}\cdot\frac{\omega}{\rho\omega_s^2}\mathrm{j} & \mathbf{I}\left(1 - \frac{\omega_t^2 - \omega^2}{\omega_b^2}\right) \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix}. \tag{51}$$

Note that the $\tilde{\mathbf{F}}_b$ and $\tilde{\mathbf{V}}_b$ are different from those given earlier. Since this partial system is not symmetric, then in the above equation

$$\tilde{\mathbf{M}}^{FF} \neq \tilde{\mathbf{M}}^{VV}. \tag{52}$$

However, the property illustrated in Eq. (48) is still valid. In order to compute the output quantities, the matrix in Eq. (51) is inverted to obtain

$$\begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{V}}_b \end{Bmatrix} = \begin{bmatrix} \mathbf{I}\left(1 + \frac{\omega^2}{\omega_s^2}\right) & -\mathbf{I}\left[\frac{\rho_b(\omega^2 - \omega_t^2)}{\omega^2}\frac{\omega^2 + \omega_s^2}{\omega_s^2} + \rho\right]\omega\cdot\mathrm{j} \\ \mathbf{I}\frac{\omega}{\rho\omega_s^2}\mathrm{j} & \mathbf{I}\left(1 - \frac{\omega_t^2 - \omega^2}{\omega_b^2}\right) \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{V}}_a \end{Bmatrix}. \tag{53}$$

Equation (53) illustrates that $\tilde{\mathbf{F}}_b$ will increase when the excitation frequency increases if $\tilde{\mathbf{F}}_a$ and $\tilde{\mathbf{V}}_a$ are kept constant. To that end, increased forcing frequencies lead to larger interstructure forces.

Before concluding this case study, it should be noted that the choice of the basis functions for the four-plane method has a direct impact on the matrix conditioning. In this example just presented, the basis functions were selected to be the full modal functions of the elastic beams, since the interstructural forces extend over the beam. This selection results in diagonal submatrices $\tilde{\mathbf{M}}$, which shows that the different input and output basis functions are decoupled. However, when the forcing region is small compared with the structure, a set of local basis functions should be used to increase computation efficiency. In the following section, local basis functions will be selected so that their impact on the $\tilde{\mathbf{M}}$ matrices can be numerically studied.

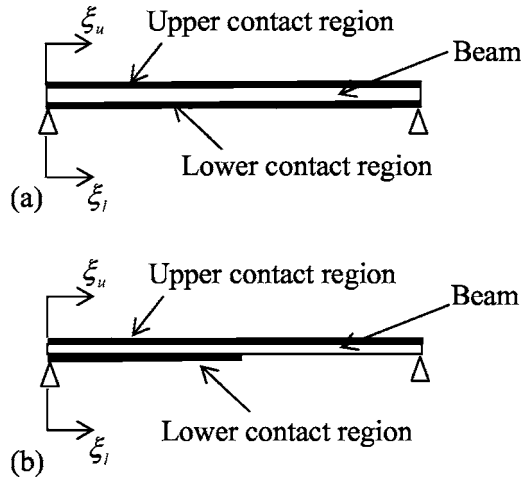


FIG. 4. Simply supported beam model. (a) Upper and lower contact region over the whole beam. (b) Upper contact region over whole beam and lower contact region over half of the beam.

E. Modal function versus local basis function selection

To examine the impact of basis function selection on the method developed here, both structures in Fig. 4 will be examined. The system illustrated in Fig. 4(a) shows a simply supported beam with an upper and lower contact region extending over the beam surface. The geometry and material properties of the beam to be considered in the study are listed in Table I. The first several computed natural frequencies of the beam are listed in Table II using the equations given by Cremer and Heckl.[5] Since the upper and lower contact region extends over the whole beam ($0 \leqslant \xi_u, \xi_l \leqslant 1$), it is suitable to select the beam modal functions (normalized sine functions) as the basis functions for both contact regions. Additionally, suppose that one is interested in the properties of the beam at a fixed frequency 159.2 Hz, which is between the first and second natural frequency. The forces and velocities acting over the upper and lower contact regions can be approximated using two terms with $L=1$, $N=2$ in Eq. (1), and the full $[\tilde{\mathbf{M}}]_a$ matrix is computed using Eqs. (35) and (36). Thus

TABLE I. The geometry and material properties of a beam.

Young's modulus, E (GPa)	206.8
Density, ρ_s(kg/m^3)	7820
Beam span, L (m)	1
Beam thickness (m)	0.05
Beam width (m)	0.05

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{S}}_a \end{Bmatrix}_{\mathrm{A}} = \begin{bmatrix} 1 & 0 & 9.0\times 10^6 & 0 \\ 0 & 1 & 0 & -1.5\times 10^7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{S}}_b \end{Bmatrix}_{\mathrm{A}} = [\mathbf{M}]_{\mathrm{A}} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{S}}_b \end{Bmatrix}_{\mathrm{A}}, \tag{54}$$

where the subscript "A" denotes the structure in Fig. 4(a), $\tilde{\mathbf{S}}_a = \tilde{\mathbf{V}}_a/(\omega \mathrm{j})$ and $\tilde{\mathbf{S}}_b = \tilde{\mathbf{V}}_b/(\omega \mathrm{j})$. Examination of $[\mathbf{M}]_{\mathrm{A}}$ also demonstrates that the determinant of $[\mathbf{M}]_{\mathrm{A}}$ is 1 and the second norm condition number[10] of $[\mathbf{M}]_{\mathrm{A}}$ is 2.2×10^{16}. The high condition number is due to the fact that $\tilde{\mathbf{M}}^{FV}$ has off-diagonal terms that are much larger than the diagonal terms. Also note that by using a two-term expansion in Eq. (1), a truncation of the series expansion is employed as only the first two basis functions are of interest at this relatively low excitation frequencies.

TABLE II. Select computed natural frequencies of the beam in Fig. 4.

Mode number	1	2	3	4	5	6	7	8	9
Frequency (Hz)	117	466	1049	1866	2915	4197	5713	7462	9444

Consider now the system illustrated in Fig. 4(b), which involves a simply supported beam with the upper contact region over the whole beam ($0 \leqslant \xi_u \leqslant 1$) and the lower contact region only over half of the beam ($0 \leqslant \xi_l \leqslant 0.5$). In this example, the forces and velocities over the upper contact region are represented with normalized modal functions that are the same as those in the previous example. The forces and velocities over the lower contact region, however, are represented with local normalized sine functions

$$\phi_i(x) = \begin{cases} 2\sin(2\cdot i\cdot \pi\cdot \xi_l) & 0 \leqslant \xi_l \leqslant 0.5 \\ 0 & \text{otherwise.} \end{cases} \tag{55}$$

This expansion is simply a Fourier sine expansion. Suppose all the other conditions are the same as in the previous example. The four-plane matrix can be computed using Eqs. (17)–(20) and Eq. (24) as

$$\begin{Bmatrix} \tilde{\mathbf{F}}_a \\ \tilde{\mathbf{S}}_a \end{Bmatrix}_{\mathrm{B}} = \begin{bmatrix} 0.6 & -0.24 & 0 & -3.8\times 10^7 \\ 0.7 & 0 & -2.1\times 10^8 & -5.2\times 10^8 \\ 0 & 0 & 0 & -4.2 \\ 0 & 0 & 1.4 & 3.5 \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{S}}_b \end{Bmatrix}_{\mathrm{B}} = [\mathbf{M}]_{\mathrm{B}} \begin{Bmatrix} \tilde{\mathbf{F}}_b \\ \tilde{\mathbf{S}}_b \end{Bmatrix}_{\mathrm{B}}, \tag{56}$$

where the subscript "B" denotes the structure in Fig. 4(b). Examination of $[\mathbf{M}]_{\mathrm{B}}$ shows that the determinant of $[\mathbf{M}]_{\mathrm{B}}$ is 1 and the second norm condition number[10] of $[\mathbf{M}]_{\mathrm{B}}$ is 3.2×10^{17}. It is clear that $[\mathbf{M}]_{\mathrm{B}}$ in Eq. (54) has more off-diagonal terms and the condition of the matrix degrades compared with $[\mathbf{M}]_{\mathrm{B}}$ in Eq. (52). The non-zero off-diagonal terms are due to the fact that local basis functions in Eq. (53) are selected and these terms illustrate the important coupling between different input and output basis functions. At this point, it is worth noting that for some cases there is a possible ill-conditioning problem like that present in other substructuring methods.[11] This ill-conditioning can lead to computational difficulties. However, this problem can be effectively resolved using numerical methods like truncated singular value decomposition.[11] The numerical recipes are well documented in existing literature.[12] Therefore, the application of these methods will not be discussed here.

In Eqs. (52) and (54), the forces and responses are represented using a two term expansion. For these cases, the natural frequencies are well-separated and only two modes dominate the response at 159.2 Hz. Therefore, a two term expansion seems reasonable for characterizing the structures if spatial variations of low physical wave-number quantities are of a primary interest. For other cases and other frequencies, more terms might be needed to fully characterize the structures. It is also worth noting that in some cases the determinant of the $[\mathbf{M}]$ matrix might not be unity if truncation errors cannot be neglected. Requirements for selecting the appropriate number of the terms should be considered in further research.

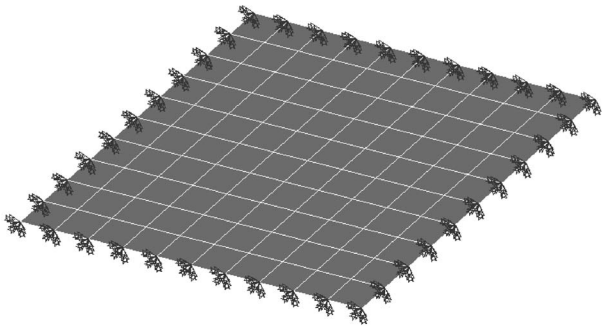


FIG. 5. FE model of a sheet metal part.

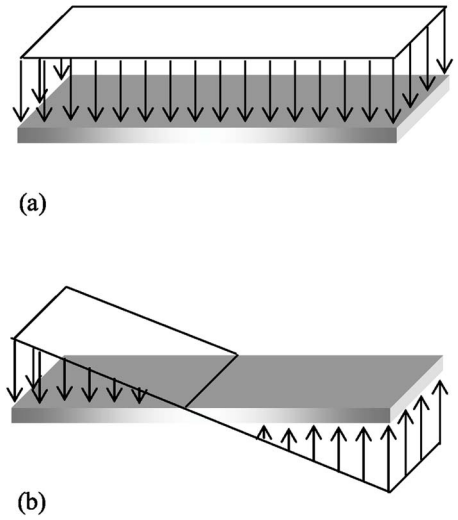


FIG. 6. (a) Uniformly distributed spatial force, (b) linearly varying momentlike spatial force.

In the following section, the four-plane method will be applied to a two-dimensional structure. The coupling between the input and output force and velocity distributions will be illustrated as functions of the Helmholtz number. Additionally, the example shows the potential for applying the four-plane method to other structures besides the vibration isolators considered above.

F. Application of the four-plane method to a two-dimensional structure

The purpose of this example is to provide an illustration in which the more commonly available finite element method is used to formulate the four-plane model. Illustrated in Fig. 5 is a FE model of a two-dimensional structure, which is a large piece of sheet metal. This case might represent, for example, a flat structure connected to other structures via two square vibration isolators. The structure was meshed with linear shell elements with a thickness h_p of 0.9 mm, a

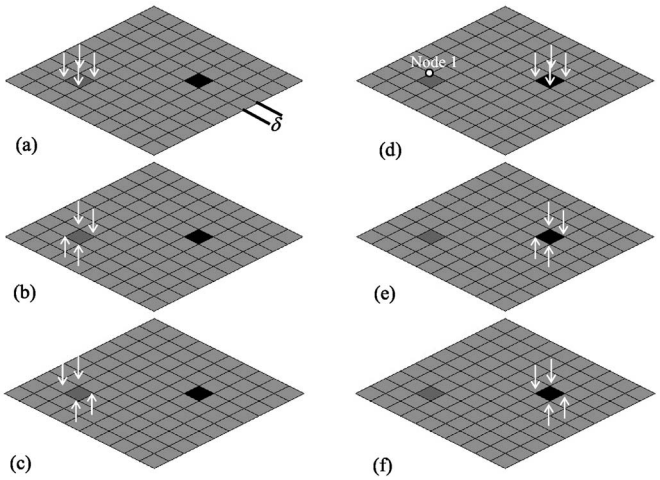


FIG. 7. Basis functions over the input region (a) first, (b) second, (c) third, and output region (e) first, (f) second, and (g) third.

Young's modulus E_p of 2e5 N/mm^2, a density ρ_p of 7.8e-9 kg/mm^3, and Poisson's ratio ν_p of 0.3. The size of this sheet is 1000 mm$\times$1000 mm and it is clamped around its edges with all six DOF (degrees of freedom) constrained. Note that the structure is a modal-rich part with the first natural frequency around 7 Hz. In this example, a uniformly distributed spatial force and a momentlike spatial force are of interest. Thus the basis functions are chosen as a uniformly distributed force and a linearly varying force across the input and output regions as illustrated in Fig. 6. The basis functions are illustrated as nodal quantities over the input and output regions in Fig. 7. Note that the input region is denoted by the light gray area, whereas the output region is the dark area. The basis forces are applied at the corner nodes of these regions to approximate a uniformly distributed spatial force and two momentlike spatial forces for the input region (denoted as the "a" region) and output regions (denoted as the "b" region), respectively. In order to apply the four-plane method, first an mobility matrix like that in Eq. (15) needs to be obtained

$$\begin{Bmatrix} \tilde{\mathbf{V}}_{a1} \\ \tilde{\mathbf{V}}_{a2} \\ \tilde{\mathbf{V}}_{a3} \\ \tilde{\mathbf{V}}_{b1} \\ \tilde{\mathbf{V}}_{b2} \\ \tilde{\mathbf{V}}_{b3} \end{Bmatrix} = \begin{bmatrix} \tilde{\mathbf{A}}_{11} & \tilde{\mathbf{A}}_{12} & \tilde{\mathbf{A}}_{13} & \tilde{\mathbf{B}}_{11} & \tilde{\mathbf{B}}_{12} & \tilde{\mathbf{B}}_{13} \\ \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} & \tilde{\mathbf{A}}_{23} & \tilde{\mathbf{B}}_{21} & \tilde{\mathbf{B}}_{12} & \tilde{\mathbf{B}}_{23} \\ \tilde{\mathbf{A}}_{31} & \tilde{\mathbf{A}}_{32} & \tilde{\mathbf{A}}_{33} & \tilde{\mathbf{B}}_{31} & \tilde{\mathbf{B}}_{32} & \tilde{\mathbf{B}}_{33} \\ \tilde{\mathbf{C}}_{11} & \tilde{\mathbf{C}}_{12} & \tilde{\mathbf{C}}_{13} & \tilde{\mathbf{D}}_{11} & \tilde{\mathbf{D}}_{12} & \tilde{\mathbf{D}}_{13} \\ \tilde{\mathbf{C}}_{21} & \tilde{\mathbf{C}}_{12} & \tilde{\mathbf{C}}_{23} & \tilde{\mathbf{D}}_{11} & \tilde{\mathbf{D}}_{12} & \tilde{\mathbf{D}}_{23} \\ \tilde{\mathbf{C}}_{31} & \tilde{\mathbf{C}}_{32} & \tilde{\mathbf{C}}_{33} & \tilde{\mathbf{D}}_{31} & \tilde{\mathbf{D}}_{32} & \tilde{\mathbf{D}}_{33} \end{bmatrix} \begin{Bmatrix} \tilde{\mathbf{F}}_{a1} \\ \tilde{\mathbf{F}}_{a2} \\ \tilde{\mathbf{F}}_{a3} \\ \tilde{\mathbf{F}}_{b1} \\ \tilde{\mathbf{F}}_{b2} \\ \tilde{\mathbf{F}}_{b3} \end{Bmatrix}. \tag{57}$$

The computation of this mobility matrix can be illustrated by the computation of one entry $\tilde{\mathbf{A}}_{11}$. First, the basis force f_{a1} with a unit amplitude is applied at the input region. The responses ν are then computed and mapped to the basis function f_{a1} to give

$$\tilde{\mathbf{A}}_{11} = \sum_i f_{a1}(i) \cdot \nu(i), \tag{58}$$

where $f_{a1}(i)$ and $\nu(i)$ are the basis function and velocity at node i. Each entry in the mobility matrix is obtained using

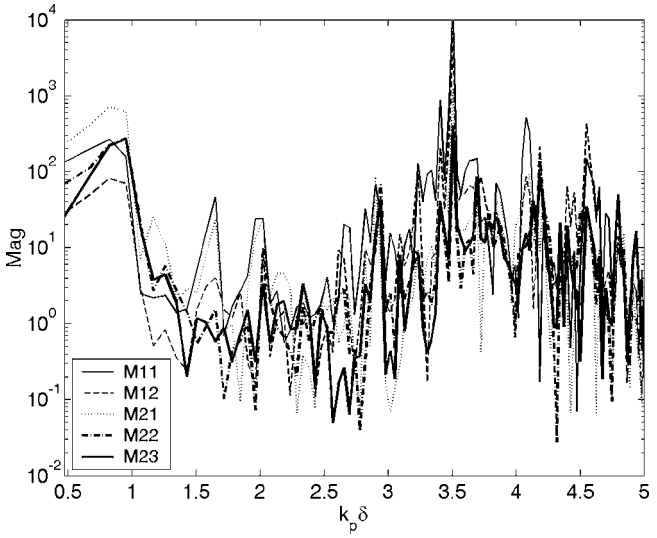


FIG. 8. The values of entries of $\tilde{\mathbf{M}}^{FF}$ as functions of Helmholtz number $k_p\delta$.

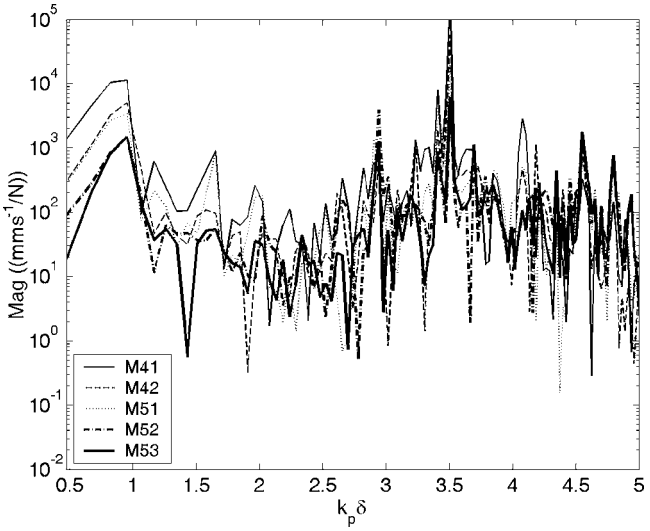


FIG. 10. The values of entries of $\tilde{\mathbf{M}}^{VF}$ as functions of Helmholtz number $k_p\delta$.

the same method programmed in the *Post26* module in ANSYS. Then matrix $[\tilde{\mathbf{M}}]_{\text{sheet}}$ for the sheet metal can be computed using Eq. (24). Matrix $[\tilde{\mathbf{M}}]_{\text{sheet}}$ has a dimension of 6×6. Figures 8–11 illustrate the entries of $[\tilde{\mathbf{M}}]^{FF}$, $[\tilde{\mathbf{M}}]^{FV}$, $[\tilde{\mathbf{M}}]^{VF}$, and $[\tilde{\mathbf{M}}]^{VV}$ as functions of the Helmholtz number, respectively. Two items are worth noting from these figures. First, the imaginary parts for all the terms in $[\tilde{\mathbf{M}}]^{FF}$ and $[\tilde{\mathbf{M}}]^{VV}$ as well as the real parts for all the terms in $[\tilde{\mathbf{M}}]^{FV}$ and $[\tilde{\mathbf{M}}]^{VF}$ are zeros due to the fact that there is no damping for this example. Second, the five values shown in each figure completely describe each sub-$\tilde{\mathbf{M}}$ matrix due to the repeated values of some of the elements as shown in Table III. In these figures, k_p is the wave number given by[13]

$$k_p = \sqrt{\omega}\left[\frac{\rho_p h_p}{D_p}\right]^{1/4}, \tag{59}$$

where D_p is the plate bending stiffness given by

$$D_p = \frac{E_p h_p^3}{12(1-\nu_p^2)}. \tag{60}$$

δ is the contact length illustrated in Fig. 7. Here $\tilde{\mathbf{M}}^{FF}$ represents the reciprocal of the spatial force transmissibility when the output region is blocked and $\tilde{\mathbf{M}}^{VV}$ represents the reciprocal spatial velocity transmissibility when the output is free. $\tilde{\mathbf{M}}^{FV}$ represents the impedance between input forces and output velocities when the output is free, while $\tilde{\mathbf{M}}^{VF}$ represents the mobility between input velocities and output forces when the output is blocked. Recall that a large coupling between the basis functions would be indicated by large values for the coefficients found using Eqs. (17)–(20). For example, a large coupling between the velocity and force basis functions would results in a larger C_{jk} in Eq. (17). Therefore, a smaller $\tilde{\mathbf{M}}$ entry value indicates a stronger coupling between corresponding basis functions, as the $\tilde{\mathbf{M}}^{FV}$ matrix involves an in-

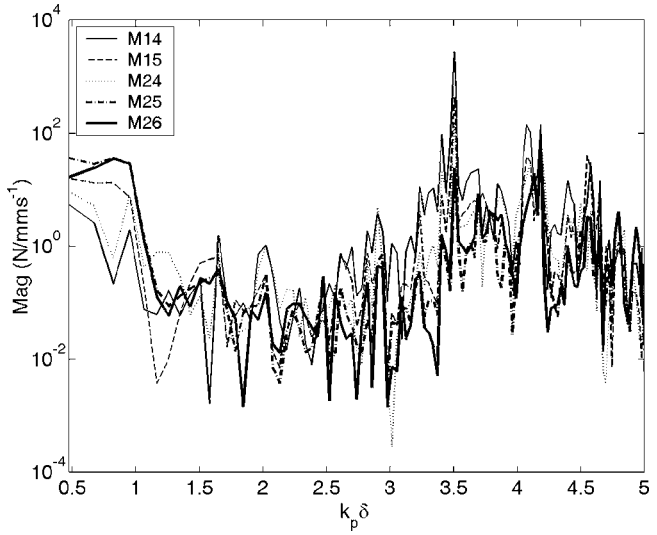


FIG. 9. The values of entries of $\tilde{\mathbf{M}}^{FV}$ as functions of Helmholtz number $k_p\delta$.

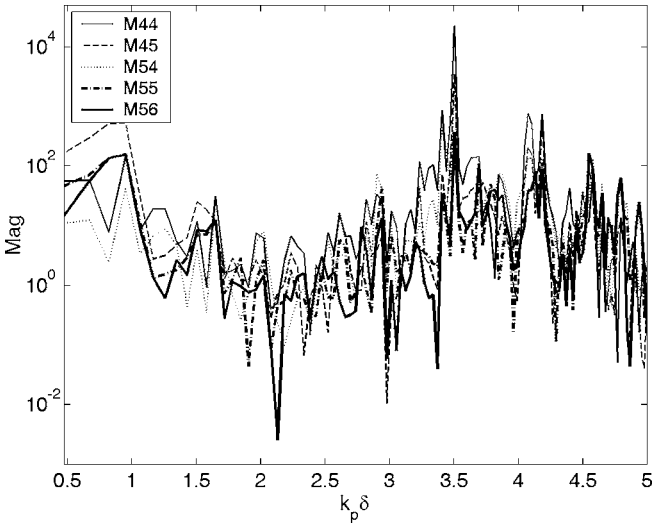


FIG. 11. The values of entries of $\tilde{\mathbf{M}}^{VV}$ as functions of Helmholtz number $k_p\delta$.

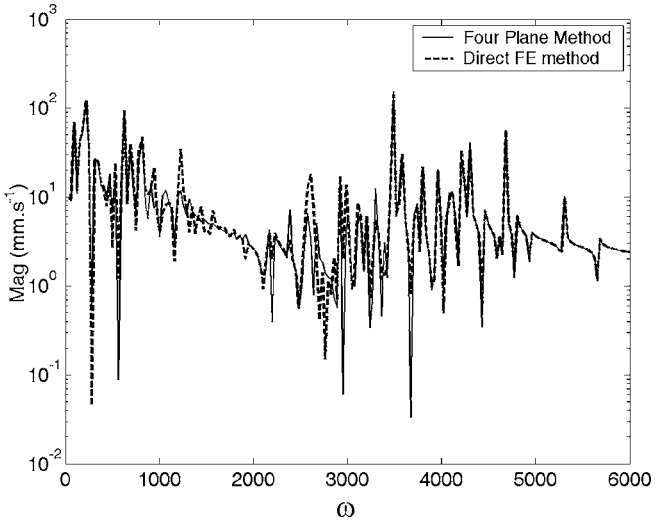


FIG. 12. Computed response at Node 1 using the four-plane method (—) and the direct FE method (- - -).

version in Eq. (26). This can also be explained by the fact that each entry of the $\tilde{\mathbf{M}}$ matrix represents a ratio between input and output as indicated in Eq. (5). Consequently, a small ratio produces a large output when the input is kept constant.

It is interesting to note that dips or peaks in these figures occur at the resonances where the basis functions more closely match the modal functions over the contact region. Illustrated in Figs. 9 and 11 are the entries of $\tilde{\mathbf{M}}^{FV}$ and $\tilde{\mathbf{M}}^{VV}$ when the output region is free. The values of $\tilde{\mathbf{M}}^{FV}$ and $\tilde{\mathbf{M}}^{VV}$ when the output region is blocked are illustrated in Figs. 8 and 10. For the case where $k_p\delta \approx \pi/4$, the scale of the contact region is approximately one-eighth of the structural wavelength and the corresponding frequency is 15 Hz. The dips of $\tilde{\mathbf{M}}_{14}$ and $\tilde{\mathbf{M}}_{44}$ indicate that the coupling between uniformly distributed basis functions dominate at that frequency. This is due to the fact that the dominant mode at this frequency is a rocking mode (second mode of the structure) where the mode shape over the input and output more closely matches the uniformly distributed basis function. This conclusion can also apply to other dips in these figures.

For the cases where $\pi/4 < k_p\delta < 8\pi/5$, the scale of the contact region is between one-eighth of the structural wavelength and four-fifths of the structural wavelength. It is clear that the momentlike spatial functions (thicker curves) dominate at most frequencies indicating a stronger coupling between the momentlike forces than for the uniformly distributed forces. These results follow the general rule that moments cannot be neglected when the structural wavelength is comparable to the size of the contact region.[6,14] While this example is somewhat simplified, it does demonstrate some of the insight that can be gained by using the approach.

Once the matrix $[\tilde{\mathbf{M}}]_{\text{sheet}}$ is known, the input forces and velocities can be computed from measurements across the output region. Suppose the output region is free and the output velocities are measured. Equation (5) can then be used to determine the velocities across the input region. Figure 12 shows the computed response at node 1 using both the four-plane method and conventional FE method. Note the location of node 1 is labeled in Fig. 7. The results from the four-plane method match the direct solution from the FE method pretty well over most of the frequency range. The deviations at some frequencies are in part due to truncation errors associated with the use of a series representation.

TABLE III. M matrix for the 2D plate structure.

$\tilde{M}_{11}$	$\tilde{M}_{12}$	$\tilde{M}_{13}=\tilde{M}_{12}$	$\tilde{M}_{14}$	$\tilde{M}_{15}$	$\tilde{M}_{16}=-\tilde{M}_{15}$
$\tilde{M}_{21}$	$\tilde{M}_{22}$	$\tilde{M}_{23}$	$\tilde{M}_{24}$	$\tilde{M}_{25}$	$\tilde{M}_{26}$
$\tilde{M}_{31}=\tilde{M}_{21}$	$\tilde{M}_{32}=\tilde{M}_{23}$	$\tilde{M}_{33}=\tilde{M}_{22}$	$\tilde{M}_{34}=\tilde{M}_{24}$	$\tilde{M}_{35}=-\tilde{M}_{26}$	$\tilde{M}_{36}=-\tilde{M}_{25}$
$\tilde{M}_{41}$	$\tilde{M}_{42}$	$\tilde{M}_{43}=\tilde{M}_{42}$	$\tilde{M}_{44}$	$\tilde{M}_{45}$	$\tilde{M}_{46}=-\tilde{M}_{45}$
$\tilde{M}_{51}$	$\tilde{M}_{52}$	$\tilde{M}_{53}$	$\tilde{M}_{54}$	$\tilde{M}_{55}$	$\tilde{M}_{56}$
$\tilde{M}_{61}=\tilde{M}_{51}$	$\tilde{M}_{62}=\tilde{M}_{53}$	$\tilde{M}_{63}=\tilde{M}_{52}$	$\tilde{M}_{64}=\tilde{M}_{54}$	$\tilde{M}_{65}=-\tilde{M}_{56}$	$\tilde{M}_{66}=-\tilde{M}_{55}$

V. CONCLUSIONS

This work presents the development of a method that can be used to characterize the vibratory response of a structure at frequencies where the wavelength is comparable to the size of the excitation contact region. This method is based on an extension of the traditional four-pole method. Instead of defining the quantities at points, or poles, the spatial quantities are described over planes of contact. Consequently, the method is referred to as the four-plane method. In using the same approach as for the well-developed and widely accepted four-pole method, some of the benefits of that method extend to the method developed here. For example, a complex system can be easily assembled using building block techniques. To demonstrate the use of this method, a model for a simplified vibration isolator was constructed and some physical insight into the response was gained. In order to demonstrate the impact of the choice of different basis functions on the $\mathbf{M}$ matrix, two structures were studied. It was shown that when local basis functions are selected, the condition of the $\mathbf{M}$ matrix will degrade. However, the difficulties can be resolved using some numerical methods. Finally, an FE-based four-plane method was applied to a clamped sheet metal structure and the coupling effects between the spatial input and output quantities were examined. It was shown that when the scale of the contact area is comparable to the wavelength, the coupling effect between the momentlike forces is much higher than the uniformly distributed forces, which agrees with previous studies of similar structures.[6,14] The results of this work will be extended to experimental approaches and more complicated structures in future research.

ACKNOWLEDGMENT

The authors acknowledge the support of this work by the Center for Advanced Vehicle Technologies at the University of Alabama.

[1]C. T. Molloy, "Use of four-pole parameters in vibration calculations," J. Acoust. Soc. Am. **29**, 842–853 (1957).
[2]J. C. Snowdon, "Mechanical four-pole parameters and their application," J. Sound Vib. **15**, 307–323 (1971).

[3]Acoustics and vibration - laboratory measurement of vibro-acoustic transfer properties of resilient element. Part 1. ISO 10846-1 (1997).
[4]Acoustics and vibration - laboratory measurement of vibro-acoustic transfer properties of resilient element. Part 2. ISO 10846-2 (1997).
[5]L. Cremer and M. Heckl, *Structure-borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies* (Springer Verlag, New York, 1988).
[6]C. Norword, "Surface mobility of a circular contact area on an infinite plate," J. Sound Vib. **202**, 95–108 (1997).
[7]A. J. Keane and W. G. Price, *Statistical Energy Analysis: An Overview, With Applications in Structural Dynamics* (Cambridge University Press, Cambridge, England, 1997).
[8]H. G. D. Goyder and R. G. White, "Vibrational power flow from machines into built-up structures, Part III: Power flow through isolation systems," J. Sound Vib. **68**, 97–117 (1980).
[9]Y. Liu and W. Steve Shepard, Jr., "An improved spatial operational force reconstruction method ," in *Proc. SECTAM XXII*, Tuskegee, AL, 2004, pp. 314–322.
[10]G. H. Golub, *Matrix Computation* (Johns Hopkins University Press, Baltimore, Maryland, 1996), pp. 52–56.
[11]T. C. Lim and J. Li, "A theoretical and computational study of the FRF-based substructuring technique applying enhanced Least Square and TSVD approaches," J. Sound Vib. **231**(4), pp. 1135–1157 (2000).
[12]P. C. Hansen, *Rank-deficient and Discrete Ill-posed Problems* (SIAM, Philadelphia, 1998).
[13]J. F. Doyle, *Wave Propagation in Structures* (Springer-Verlag, New York, 1997).
[14]M. A. Sanderson, "Vibration isolation: Moments and rotations included," J. Sound Vib. **198**(2), pp. 171–191 (1996).

Longitudinal vibration frequencies of steadily whirling rods[a)]

W. S. Shum[b)] and R. D. Entwistle
Department of Mechanical Engineering, Curtin University of Technology, WA, Australia

(Received 7 May 2005; revised 18 October 2005; accepted 7 November 2005)

Past researchers suggested that "static instabilities" exist at certain rotational speeds of whirling rods. This paper shows these instabilities are an artefact of the material constitutive laws that are being used well outside their range of applicability. An alternative approach is developed where strains due to rotation are separated from the superimposed vibration. This enables the generally predicted lowering of longitudinal natural frequencies with rotational speed shown to be simply a result of the bulk changes in the geometry of whirling rods. Steady-state equations of whirling rods are formulated in Lagrangian coordinates. Due to the nonlinear nature of the governing equations, an original numerical method is applied to solve the problem. Numerical results are compared with analytical results obtained from the linearized uniaxial model. There is close agreement between these two models at low angular velocities. However, at high angular velocities, discrepancies between them arise, confirming that the nonlinear strain-displacement relationship has significant effect on the results and the inferred "static instabilities." This approach first solves the "static" problem of the deformed geometry of a highly strained whirling rod before longitudinal natural modes are determined by classical methods. Furthermore, conditions for existence and uniqueness of solutions are derived. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146086]

PACS number(s): 43.40.Cw, 43.40.Ga [JGM] Pages: 909–916

I. INTRODUCTION

The longitudinal, axial, or extensional frequencies of a rod rotating about one end are of interest in the design of turbine and helicopter blades. Methods of estimating the frequencies have been published since 1963. The problem remains unresolved, as disagreement ("controversy," Venkatesan and Nagaraj[1]) exists on whether the frequencies are increasing or decreasing with increasing rotational speed. Furthermore, it is generally agreed that static longitudinal strain becomes unbounded when a certain rotational speed is reached. This phenomenon has been called a "static resonance" or "instability." Chronologically, the work has been developing as follows. Bhuta and Jones[2] were exploring problems involving small strains and where the use of Lagrangian coordinates (undeformed) yielded improved models compared to the use of Eulerian (deformed) coordinates. While they identified the possibility of a critical speed, it followed the assumption of small strains but did not recognize that the strains near critical speeds were well beyond being "small" (see Ref. 3). As Brunelle[4] did not cite Bhuta and Jones,[2] he appears to have discovered this instability independently. Anderson[5] reconsidered the problem by incorporating the nonlinear strain equations and separated the analysis by superimposing the "initial stress" (static equilibrium) problem onto the "deformed motion" (dynamic) problem of linearized, small amplitude vibrations. Although such an assumption for nonlinear equations did not seem to be appealing, he asserted that the longitudinal natural frequencies increased with rotational speed and excluded the possibility of static instability. Hodges'[6] analysis suggested that the sign of the frequency shift with rotation depended on the hardening or softening characteristic of the strain-displacement property of the material. Venkatesan and Nagaraj[1] claimed that Hodges'[6] equations contained an error and suggested that premature ordering "completely changed" the characteristic of the behavior of the longitudinal natural frequencies with rotational speed. Hodges[7] refuted these conclusions by maintaining that an unknown material parameter (the coefficient of the nonlinear strain term) governed the response. A decade later Hodges and Bless[3] further reinforced the earlier findings by proposing simple nonlinear strain energy models that included a cubic term in the displacement gradient. They showed that instability existed only when the coefficient of this term was less than one, signifying a softening stiffness. They noted that instability only occurred at strains well beyond the range of elastic action for most materials. The question of longitudinal natural frequencies was not addressed. O'Reilly and Turcotte[8] approached the problem using the theory of a Cosserat rod and superimposed the small vibrational deformations on large static deformations using a nonlinear rod theory. However, they left the specific form of the strain energy function unspecified and no solutions were offered. None of the models outlined above had included the Poisson's effect of lateral contraction of the bar. However, in O'Reilly and Turcotte's later paper with Kindaid,[9] they did include lateral contraction of the bar and showed that instability coincided with the condition of the cross-sectional area of the bar somewhere reducing to zero. They concluded that this lateral contraction had a significant influence on the displacement. Some numerical solutions were offered where a whirling bar is fixed at one end. To enforce this boundary condition they also constrained the

[a)]Portions of this work were presented in "On Axial Natural Frequencies of Whirling Rods," Proceedings of the 4th Australasian Congress on Applied Mechanics, Melbourne, Australia, 16–18 February 2005, pp. 219–224.

[b)]Electronic mail: waisun.shum@student.curtin.edu.au

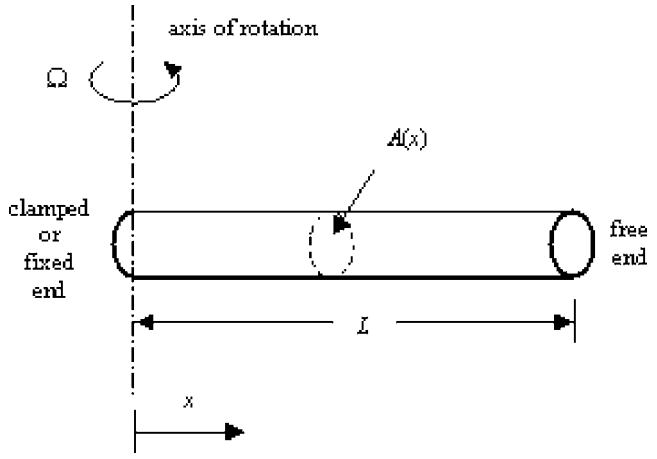


FIG. 1. Geometric configuration in Lagrangian coordinate x of a rod statically rotating about the axis of rotation at an angular velocity Ω. $A(x)$ is the cross-sectional area of the rod. L is the original length of the nonrotating undeformed rod.

lateral contraction at the axis of rotation, leading to a lateral strain profile that they termed a "boundary layer." This paper offers an alternative approach. Firstly, the problem of estimating the longitudinal vibration frequencies is to superimpose the small amplitude vibration problem onto the static (equilibrium) deformation problem. Secondly, the effects of lateral contraction are included with consequent effects on density and mass distribution. An original numerical method is applied to the static problem. The method can handle nonlinear strain-displacement relationships. The natural frequencies are then calculated using a standard substructuring (receptance) method. It is proposed that at moderate strains, the static extension of the bar by longitudinal strains is the primary cause of the lowering of longitudinal natural frequencies.

II. GOVERNING EQUATIONS

Some previous work[2] had applied the free end boundary condition of whirling rods in Eulerian coordinates. But this free end position in Eulerian coordinates is exactly the same as the free end of the rod when it is not rotating. However, this is obviously not true, because when the rod is rotating, that free end in Eulerian coordinates under axial strain must be different from the free end position of the nonrotating rod. Therefore, problems formulated in Lagrangian coordinates are better modelled than those in Eulerian coordinates because of the proper implementation of the free end boundary condition. Formulation of steady whirling rods governing equations in Lagrangian coordinates will be performed in this paper. Geometric configuration of a prismatic static whirling rod is shown in Fig. 1.

A. Linear uniaxial model

Previously, it had been shown that at low angular velocity Ω, rotational motion of the rod at steady state could be described in Lagrangian coordinate x by the linear uniaxial model with clamped-free boundary conditions,[2]

$$E\frac{d^2u}{dx^2} = -\rho_0(x+u)\Omega^2, \tag{1a}$$

$$u(0) = 0, \tag{1b}$$

$$\left[\frac{du}{dx}\right]_{x=L} = 0, \tag{1c}$$

where $0<x<L$, E is Young's modulus or the modulus of elasticity, ρ_0 is the material density of a nonrotating undeformed rod, $u(x)$ is the longitudinal, axial, or extensional displacement at x, and L is the length of a nonrotating undeformed rod. Equations (1) assume small strain analysis with a linear strain-displacement relationship. In addition, Poisson's effect is neglected by keeping the material density and cross-sectional area constant. The analytical solution for Eqs. (1) is[2]

$$u(x) = \frac{\sin(kx)}{k\cos(kL)} - x, \quad k = \sqrt{\frac{\rho_0}{E}}\Omega.$$

It is apparent from the above solution that $u(x)$ does not exist when $\Omega = \sqrt{[(2N-1)\pi/2L]^2 E/\rho_0}$, for $N=1,2,3,\ldots$. This means that the solution obtained in Lagrangian coordinates does not exist at certain angular velocities Ω. However, before reaching the lowest angular velocity Ω when $N=1$, the material could already fail at yield point as the strains that occur at these angular velocities are so large that the small strain analysis for the linear uniaxial model becomes invalid. As a result, this model is limited to small strains as no allowance is made for the reduction in $A(x)$ due to transverse Poisson's strain. Also, changes in $A(x)$ due to Poisson's strain necessarily imply density redistribution. Hence, the linear uniaxial model is not representative of the situation at larger strains due to higher angular velocity Ω.

B. Proposed nonlinear model

Assuming the rod material is isotropic with only axial normal stress $\sigma(z)$ present, then the one-dimensional Hooke's law in Eulerian coordinates can be written as $E\varepsilon(z)=\sigma(z)$, where z is the Eulerian coordinate and $\varepsilon(z)$ is axial normal strain at z of a deformed rod. But from physical law, $\sigma(z)=F(z)/A(z)$, where $F(z)$ is the force distribution at z along the deformed rod and $A(z)$ is cross-sectional area of a rod at z of a deformed rod. Substituting the expression for $\sigma(z)$ into the 1-D Hooke's law,

$$F(z) = EA(z)\varepsilon(z), \tag{2}$$

but the centripetal force

$$F(z) = \int_{x+u(x)}^{L+u(L)} \rho(z)A(z)z\Omega^2\,dz, \tag{3}$$

where $z=x+u(x)$, $\rho(z)$ is the material density at z of a deformed rod. Substituting Eq. (3) into Eq. (2),

$$EA(z)\varepsilon(z) = \int_{x+u(L)}^{L+u(L)} \rho(z)A(z)z\Omega^2\, dz. \tag{4}$$

In order to impose the free end boundary condition properly as mentioned above, Eq. (4) will be transformed into the Lagrangian coordinate x. By conservation of mass, $\int_0^{s+u(s)}\rho(z)A(z)\, dz=\int_0^s\rho_0A_0\, dx$, where A_0 is the uniform cross-sectional area of a nonrotating undeformed rod and s is any arbitrary position on the rod. Changing the integration w.r.t. x gives $\int_0^s(\rho(x)A(x)(dz/dx)-\rho_0A_0)\, dx=0$, where $A(x)$ is cross-sectional area of a rod at x of a deformed rod. Since this integration is true $\forall s\in[0,L]$, it must follow that $\rho(x)A(x)dz/dx=\rho_0A_0$. By applying this result and changing z to x in Eq. (4),

$$EA(x)\varepsilon(x) = \int_x^L \rho_0A_0(x+u)\Omega^2\, dx, \tag{5}$$

with $u=u(x)$ and $0\leqslant x\leqslant L$. From Poisson's relation, $\varepsilon_L(x)=-\nu\varepsilon(x)$, where ν is Poisson's ratio, $\varepsilon_L(x)$ is the lateral strain at the Lagrangian coordinate x, and $\varepsilon(x)$ is longitudinal normal strain along the Lagrangian coordinate x. Since $\varepsilon_L(x)=[d(x)-d_0]/d_0=-\nu\varepsilon(x)$, where $d(x)$ is the diameter of a cross-sectional area along Lagrangian coordinate x and d_0 is the uniform diameter of a nonrotating undeformed rod, it follows that $d(x)=d_0\,[1-\nu\varepsilon(x)]$. As a result,

$$A(x) = \frac{\pi}{4}[d(x)]^2 = A_0[1-\nu\varepsilon(x)]^2. \tag{6}$$

Putting Eq. (6) into Eq. (5),

$$E[1-\nu\varepsilon(x)]^2\varepsilon(x) = \int_x^L \rho_0(x+u)\Omega^2\, dx. \tag{7}$$

By inserting the nonlinear strain-displacement relationship,[10] $\varepsilon(x)=du/dx+\frac{1}{2}(du/dx)^2$, into Eq. (7),

$$E\left[1-\nu\left(\frac{du}{dx}+\frac{1}{2}\left(\frac{du}{dx}\right)^2\right)\right]^2\left[\frac{du}{dx}+\frac{1}{2}\left(\frac{du}{dx}\right)^2\right] = \int_x^L \rho_0(x+u)\Omega^2\, dx, \quad 0<x<L. \tag{8a}$$

For clamped-free boundary conditions,

$$u(0) = 0, \tag{8b}$$

$$\left[\frac{du}{dx}+\frac{1}{2}\left(\frac{du}{dx}\right)^2\right]_{x=L} = 0. \tag{8c}$$

Equations (8) are the steady-state integro-differential equations for axial displacement $u(x)$ in Lagrangian coordinate x that govern the strain of a whirling prismatic rod. Although this derivation is being done on a rod with circular cross section, beams with noncircular cross-sectional area can be derived in a similar fashion by including the cross-sectional area profile into the governing equation. This proposed model improves upon Eqs. (1) because Poisson's effect, which accounts the variations of material density and cross-sectional area, has been included. Similar to the linear uniaxial model, the nonlinear governing equation is only valid in the linear elastic region that is governed by the 1-D Hooke's law. Beyond the end of this linear elastic region, which is the yield point, the validity of the governing equations does not apply. This proposed model is deduced by including all three variations of axial displacement, material density, and cross-sectional area. Thus, it is suitable for large strain analysis.

III. NUMERICAL SOLUTION

Since it is very difficult, although not impossible, to obtain an analytical solution for nonlinear integro-differential equations, numerical solutions are being sought instead. Assuming $\{f_i(x)\}_{i=0}^{\infty}$ is an infinite sequence which forms a complete set of functions of an unknown variable $u(x)$, then $u(x)=\Sigma_{i=0}^{\infty}a_if_i(x)$, where $x\in[0,L]$ and a_i's are unknown constants to be determined. Since $\{f_i(x)\}_{i=0}^{\infty}$ is a complete set, given every $\delta>0$, there exists a positive integer N_δ such that $|\Sigma_{i=0}^{m}a_if_i(x)-\Sigma_{i=0}^{n}a_if_i(x)|<\delta, \forall m,n>N_\delta$. For the current studies, $\{f_i(x)\}_{i=0}^{\infty}$ is chosen to be $\{x^i\}_{i=0}^{\infty}$. Consider the following equations

$$\Psi_\Xi(u(x)) = g(x), \quad x\in\Xi, \tag{9a}$$

$$\Psi^1_{\partial\Xi}(u(x)) = g_1(x), \quad x\in\partial\Xi, \tag{9b}$$

$$\Psi^2_{\partial\Xi}(u(x)) = g_2(x), \quad x\in\partial\Xi, \tag{9c}$$

where Ψ_Ξ is a nonlinear integro-differential operator on the interior of domain x, $\Psi^1_{\partial\Xi},\Psi^2_{\partial\Xi}$ are nonlinear differential operators at the boundary of domain x, Ξ represents the interior of domain x, $\partial\Xi$ represents the boundary of domain x, $u(x)$ is an unknown function of variable x, and $g(x)$, $g_1(x)$, and $g_2(x)$ are known functions of variable x. In order to implement a numerical solution in a computer, $u(x)$ can only be approximated by a finite number of terms in the complete set $\{f_i(x)\}_{i=0}^{\infty}$. Therefore, $u(x)\approx\Sigma_{i=0}^{n}a_if_i(x)$ are approximate solutions of Eqs. (9), where n is a finite non-negative integer, f_i's are prescribed functions, and a_i's are unknown constants to be determined. By choosing $x_j\in\Xi$, $j=1,2,\ldots,n-1,x_0,x_n\in\partial\Xi$, Eqs. (9) can be approximated by

$$\Psi_\Xi\left(\sum_{i=0}^{n}a_if_i(x_j)\right) \approx g(x_j), \quad x_j\in\Xi, \quad j=1,2,n-1, \tag{10a}$$

$$\Psi^1_{\partial\Xi}\left(\sum_{i=0}^{n}a_if_i(x_0)\right) \approx g_1(x_0), \quad x_0\in\partial\Xi, \tag{10b}$$

$$\Psi^2_{\partial\Xi}\left(\sum_{i=0}^{n}a_if_i(x_n)\right) \approx g_2(x_n), \quad x_n\in\partial\Xi. \tag{10c}$$

Equations (10) can now be solved by Newton's method. Defining

$$e_j = \begin{cases} \Psi_{\partial\Xi}\left(\sum_{i=0}^{n} a_i f_i(x_0)\right) - g_1(x_0), & j=0, \\ \Psi_{\Xi}\left(\sum_{i=0}^{n} a_i f_i(x_j)\right) - g(x_j), & j=1,2,\ldots,n-1, \\ \Psi_{\partial\Xi}\left(\sum_{i=0}^{n} a_i f_i(x_n) - g_2(x_n)\right), & j=n, \end{cases}$$

then the stopping criteria for Newton's iteration is when the error tolerance $\Sigma_{j=0}^{n} e_j^2 \ll 1$. For the particular problem in Eqs. (8), it is required that $u(x) \in C^1(0,L]$, where $C^1[0,L]$ is a class of functions that are continuously differentiable $\forall x \in (0,L]$.

This numerical method resembles collocation, Galerkin, and Rayleigh-Ritz methods, although not exactly the same. All these methods assume the unknown solution of governing equations to be represented by a linear combination of prescribed functions, namely $\Sigma_{i=0}^{n} a_i f_i$, where a_i's are the unknown coefficients to be determined, f_i's are the prescribed functions, and n is a non-negative integer. In the end, a nonlinear algebraic system of equations is generated to be solved by the Newton's method for the unknowns a_i's. Collocation, Galerkin, and Rayleigh-Ritz methods belong to the family of projection methods. The prescribed functions f_i's are often called the basis functions. These basis functions could be orthogonal to each other. Some of the choices of these basis functions are the Chebyshev polynomials of first and second kind. Nonetheless, the prescribed functions embraced in the proposed numerical method are required to be a complete representation of the unknown solution in Newton's sense. Completeness of a solution in Newton's sense simply means that the numerical solutions produced by Newton's iteration method converge to the true solution. This criterion might not be satisfied if the chosen prescribed functions are not otherwise complete in the Newton's sense. For instance, the Chebyshev polynomials might not be a complete representation of the unknown solution in Newton's sense.

IV. RESULTS AND COMPARISONS

Numerical experiments are performed on a rod made of steel with arbitrarily chosen parameters. Relevant parameters are diameter $d_0=0.035$ m, density $\rho_0=7850$ kg/m^3, Young's modulus $E=200$ GPa, Poisson's ratio $\nu=0.29$, and length of the rod $L=1$ m.

A. Longitudinal displacements and longitudinal stresses

Comparisons between the analytical results of the uniaxial model in Eqs. (1) and the numerical results of the proposed nonlinear model in Eqs. (8) are plotted in Fig. 2 for axial displacements and Fig. 3 for axial stresses. Size of the series n in Eqs. (10) is chosen to be ten as the error tolerance is approximately $O(10^{-30})$ in the stopping criteria of Newton's method. For $n=20$, there are no significant differences in the results compared to $n=10$. For $n>20$, the error tolerance of $O(10^{-30})$ is not achievable due to significant round-off errors. Both models agree well at a low angular velocity

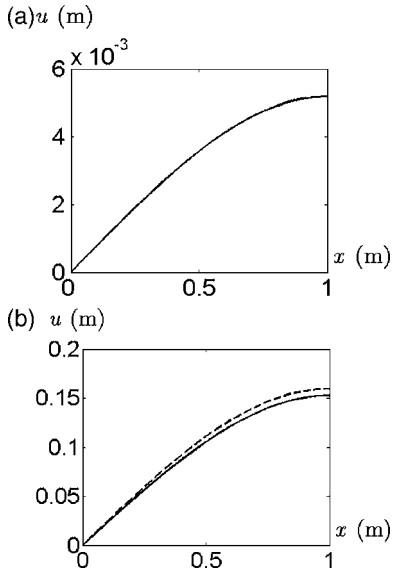


FIG. 2. Plot of longitudinal displacement $u(x)$ against Lagrangian coordinate x. Solid line represents results from Eqs. (1). Dashed line represents results from Eqs. (8). $\rho_0=7850$ kg/m^3, $E=200$ GPa, $\nu=0.29$ and $L=1$ m. (a) $\Omega=200\pi$ rad/s. (b) $\Omega=1000\pi$ rad/s.

$\Omega=200\pi$ rad/s (6000 rpm) but only qualitatively agree at high angular velocity $\Omega=1000\pi$ rad/s (30000 rpm).

Although there are computational results beyond $\Omega=200\pi$ rad/s, numerical results show that material of rods that are made of steel with the above given parameters will fail at the yield point before reaching $\Omega=200\pi$ rad/s. Thus the results in Figs. 2 and 3 are only displayed for the purpose of comparisons between the two models. Although the comparisons are beyond the yield point of the material, results show that the differences between the linear uniaxial model and the proposed nonlinear model become apparent at high

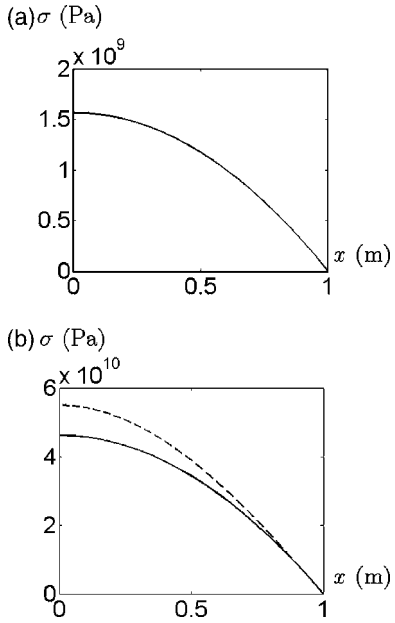


FIG. 3. Plot of longitudinal stress $\sigma(x)$ against Lagrangian coordinate x. Solid line represents results from Eqs. (1). Dashed line represents results from Eqs. (8). $\rho_0=7850$ kg/m^3, $E=200$ GPa, $\nu=0.29$, and $L=1$ m. (a) $\Omega=200\pi$ rad/s. (b) $\Omega=1000\pi$ rad/s.

TABLE I. First five longitudinal natural frequencies of the nonrotating isotropic rod.

	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$
ω_N (rad/s)	7929	23 786	39 643	55 501	71 358
f_N (rad/s)	7929	23 786	39 643	55 501	71 358

angular velocities. If a rod with a material has a linear elastic region that follows Hooke's law at high angular velocities, the proposed nonlinear model prevails. Defining $e=(\sigma_{NL}-\sigma_L)/\sigma_L$, where σ_{NL} is the axial stress predicted by Eqs. (8) at $x=0$ and σ_L is the axial stress predicted by Eqs. (1) at $x=0$, then from Fig. 3(a), $e\approx 0.39\%$ at $\Omega=200\pi$ rad/s and from Fig. 3(b), $e\approx 19\%$ at $\Omega=1000\pi$ rad/s. When $e=1\%$ and 10%, the corresponding angular velocities are at $\Omega=294\pi$ and 812π rad/s, respectively.

B. Longitudinal natural frequencies

This paper contends that the axial natural frequencies of a whirling rod depend primarily on the geometry of the deformed rod, not on the rotary motion *per se*. Hence, having determined the deformed geometry, the determination of the axial vibration frequencies can be pursued via classical methods. A classical substructuring (receptance) method is employed to compute the axial natural frequencies of the deformed rod according to the results obtained from Eqs. (8). Readers may refer to Ref. 11 for the full description of this method. In all the tables, $\omega_N=\sqrt{[(2N-1)\pi/2L]^2(E/\rho_0)}$, where $N=1,2,3,\ldots$, is the analytical axial natural frequency for non-rotating undeformed rod. Moreover, $p_N=\sqrt{[(2N-1)\pi/2L]^2(E/\rho_0)-\Omega^2}$, where $N=1,2,3,\ldots$, is the analytical axial natural frequency of the rotating deformed rod obtained from Bhuta and Jones.[2] Since small strain analysis is assumed, p_N is only applicable to rods rotating at low angular velocities. f_N, where $N=1,2,3,\ldots$, represents axial natural frequencies obtained by the classical (receptance) approach. All longitudinal natural frequencies in the tables are rounded to the nearest integer. Results and comparisons are displayed in Tables I–III.

The methods by Bhuta and Jones[2] and the classical approach both predict generally lowered axial natural frequencies if angular velocity is increased. However, the discrepancies between these two methods widen as angular velocity Ω increases further. These discrepancies are attributed to the p_N predictions using the linear model that ignores the variations of material density and cross-sectional area while the f_N predictions using the proposed nonlinear model take these two factors into account. It can be imagined that when a rod becomes longer, the wavelength of a near sinusoidal wave that is required to create a resonance is larger. If the wave speed remains constant, then the natural frequencies must be smaller accordingly. Hence the natural frequencies are decreased for an increased angular velocity of a steadily whirling rod.

TABLE II. First five longitudinal natural frequencies at angular velocity $\Omega=200\pi$ rad/s.

	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$
p_N (rad/s)	7904	23 778	39 638	55 497	71 355
f_N (rad/s)	7886	23 685	39 479	55 272	71 064

TABLE III. First five longitudinal natural frequencies at angular velocity $\Omega=1000\pi$ rad/s.

	$N=1$	$N=2$	$N=3$	$N=4$	$N=5$
p_N (rad/s)	7280	23 578	39 519	55 412	71 289
f_N (rad/s)	6713	20 858	34 857	48 836	62 808

C. Material density and diameter of the whirling rod

Although there are no analytical or experimental data to compare with, results obtained by the proposed numerical method from Eq. (8) on the material density and the diameter of the rod while it is rotating are shown in Figs. 4 and 5 respectively.

From Fig. 4, material density of the whirling rod is redistributed compared to the constant density $\rho_0=7850$ kg/m^3 of the stationary rod. Similarly from Fig. 5, the diameter of the whirling rod varies along x compared to the uniform diameter $d_0=0.035$ m of the stationary rod. In view of Fig. 5, it can be concluded that the steady whirling rod has a trumpetlike shape.

V. EXISTENCE AND UNIQUENESS OF SOLUTION

Equations (1) and (8) are termed two-point boundary value problems. Depending on the boundary conditions, these equations may have no solution, a unique solution, or more than one solution. Conditions on existence and uniqueness of solution for both linear uniaxial and the proposed nonlinear models are derived as follows.

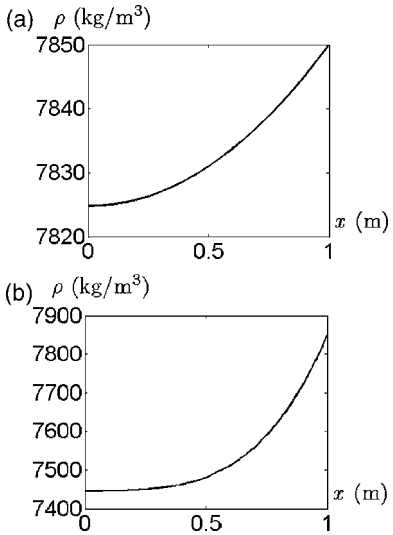


FIG. 4. Plot of density $\rho(x)$ of the rotating rod against Lagrangian coordinate x. Results are obtained by the proposed numerical method on Eqs. (8). $d_0=0.035$ m, $\rho_0=7850$ kg/m^3, $E=200$ GPa, $\nu=0.29$, and $L=1$ m. (a) $\Omega=200\pi$ rad/s. (b) $\Omega=1000\pi$ rad/s.

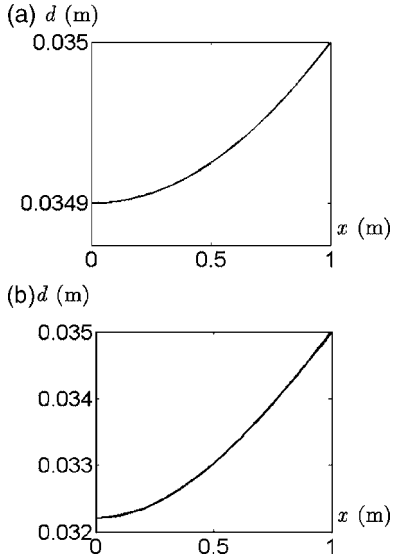


FIG. 5. Plot of diameter $d(x)$ of the rotating rod against Lagrangian coordinate x. Results are obtained by the proposed numerical method on Eqs. (8). $d_0=0.035$ m, $\rho_0=7850$ kg/m^3, $E=200$ GPa, $\nu=0.29$, and $L=1$ m. (a) $\Omega=200\pi$ rad/s. (b) $\Omega=1000\pi$ rad/s.

A. Linear uniaxial model

Theorem 1: *Consider the following inhomogeneous Helmholtz differential equations*,

$$\frac{d^2u}{dx^2}+k^2u=-k^2x, \quad 0<x<L \text{ and } k\neq 0, \tag{11a}$$

$$u(0)=0, \tag{11b}$$

$$\left[\frac{du}{dx}\right]_{x=L}=0. \tag{11c}$$

Solution of Eq. (11a) can be found to be $u(x)=c_1\cos(kx)+c_2\sin(kx)-x$. *From the boundary conditions (11b) and (11c), the arbitrary constants* c_1 *and* c_2 *can be determined from the matrix equation*

$$A\mathbf{x}=\mathbf{b}, \tag{12}$$

where

$$A=\begin{bmatrix} 1 & 0 \\ -k\sin(kL) & k\cos(kL) \end{bmatrix},$$

$$\mathbf{x}=\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad \text{and} \quad \mathbf{b}=\begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

If the rank of matrix A equals 2, then solution for Eqs. (11) exists and is unique.

Proof: (Existence). If the rank of matrix A equals 2, then $\mathbf{x}$ can be found from Eq. (12) by multiplying $\mathbf{b}$ on the left with A^{-1}. Hence solution exists.

(Uniqueness). Suppose $u_1(x)$ and $u_2(x)$ are the solutions of the equations (11) and let $\zeta(x)=u_1(x)-u_2(x)$. Then

$$\frac{d^2\zeta}{dx^2}+k^2\zeta=0, \quad 0<x<L \text{ and } k\neq 0, \tag{13a}$$

$$\zeta(0)=0, \tag{13b}$$

$$\left[\frac{d\zeta}{dx}\right]_{x=L}=0. \tag{13c}$$

Since

$$\int_0^L\left(\frac{d^2\zeta}{dx^2}+k^2\zeta\right)\frac{d\zeta}{dx}\,dx=\frac{1}{2}\int_0^L\frac{d}{dx}\left[\left(\frac{d\zeta}{dx}\right)^2+k^2\zeta^2\right]dx$$
$$=\frac{1}{2}\left[\left(\frac{d\zeta}{dx}\right)^2+k^2\zeta^2\right]_0^L,$$

by using Eqs. (13), it implies

$$k^2[\zeta(L)]^2=\left[\left(\frac{d\zeta}{dx}\right)_{x=0}\right]^2. \tag{14}$$

Now

$$\frac{d}{dx}\left[\left(\frac{d\zeta}{dx}\right)^2+k^2\zeta^2\right]=2\left(\frac{d^2\zeta}{dx^2}+k^2\zeta\right)\frac{d\zeta}{dx}=0$$

for $0<x<L$. Because of Eq. (13a), it implies

$$\left(\frac{d\zeta}{dx}\right)^2+k^2\zeta^2=c^2, \tag{15}$$

where c is a constant. If Eq. (15) is true, then it must represent an equation of an ellipse on the (x,ζ) plane with

$$\frac{d\zeta}{dx}=\alpha(x-\beta), \tag{16}$$

where α and β are constants. Integrating once,

$$\zeta=\frac{\alpha}{2}x^2-\alpha\beta x+\gamma, \tag{17}$$

where γ is a constant. Substituting Eq. (16) into Eq. (15) gives

$$\zeta^2=\frac{1}{k^2}(c^2-\alpha^2x^2+2\alpha^2\beta x-\alpha^2\beta^2). \tag{18}$$

From Eq. (17),

$$\zeta^2=\tfrac{1}{4}\alpha^2x^4-\alpha^2\beta x^3+(\alpha\gamma+\alpha^2\beta^2)x^2-2\alpha\beta\gamma x+\gamma^2. \tag{19}$$

Since Eqs. (18) and (19) must be equivalent to each other, it implies that $\alpha=0$ and $\gamma=c/k$. Therefore, $\zeta(x)=c/k$ for $0<x<L$. By enforcing Eqs. (13) and (14), $\zeta(x)=0$ for $0\leqslant x\leqslant L$. Hence the proof is completed. □

In view of Eqs. (1), solutions of Eqs. (1) do not exist when $\Omega=[(2N-1)\pi/2L]\sqrt{E/\rho_0}$, $N=1,2,3\ldots$, that is, when solutions of Eqs. (1) become singular. For $0\leqslant\Omega<(\pi/2L)\sqrt{E/\rho_0}$, solutions of Eqs. (1) exist from Sec. II A and are unique according to Theorem 1. For $(\pi/2L)\sqrt{E/\rho_0}<\Omega<(2\pi/L)\sqrt{E/\rho_0}$, solutions of Eqs. (1) are negative for some Lagrangian coordinate $x\in[0,L]$. Similar arguments can be applied to solutions when $\Omega>(2\pi/L)\sqrt{E/\rho_0}$. As a result, solution of Eqs. (1) does not exist for $\Omega\geqslant(\pi/2L)\sqrt{E/\rho_0}$. For the numerical experiment using the parameters in Sec. IV, solution of Eqs. (1) becomes invalid when $\Omega\geqslant(\pi/2L)\sqrt{E/\rho_0}\approx 7929$ rad/s. Although Bhuta and

Jones[2] did not work out this condition, they termed it the "static instabilities." Proof of Theorem 1 is a proof by contradiction which has been applied by Sagan[12] to prove the existence and uniqueness of heat and wave equations.

B. Proposed nonlinear model

Theorem 2: *Let I be the closed interval $0 \leqslant x \leqslant L$ and $u(x)$ be continuous with continuous first-order derivatives in I. If $u(x)$ satisfies Eqs. (8) and $0 \leqslant d(x) \leqslant d_0 \forall x \in I$, where $d(x)$ is the diameter of the rotating rod and d_0 is the original diameter of the nonrotating rod, then $u(x)$ is a solution of Eqs. (8). Nevertheless, $u(x)$ is not the only function in I with such properties. In addition, if $d(x) < \frac{2}{3}d_0$ for some or all $x \in I$ while the angular velocity $\Omega \neq 0$, there exists no physically correct $u(x)$ that will satisfy Eqs. (8).*

Proof: From Poisson's relation, $\varepsilon(x) = -(d-d_0)/d_0\nu$, where $d=d(x)$. Substituting this into Eq. (2) gives $d^2[d-d_0] = -(4/\pi E)d_0\nu F$, where $0 \leqslant d = d(x) \leqslant d_0$. Now let $h_1(d) = d^2[d-d_0]$. Then $h_1'(d) = 3d^2 - 2dd_0$ and $h_1''(d) = 6d - 2d_0$. When $h_1'(d) = 0$, $d = \frac{2}{3}d_0$. Hence $h_1(\frac{2}{3}d_0)$ is minimum because $h_1''(\frac{2}{3}d_0) = 2d_0 > 0$. Also, let $h_2(d) = -4/\pi E\, d_0\nu F$. Then, for each $x \in [0, L]$, one of the two intersections of $h_1(d)$ and $h_2(d)$ is the physically correct solution of Eqs. (8) at a specified Ω provided that $h_2(d) > h_1(\frac{2}{3}d_0)$, the other solution is physically meaningless, or there is only one physically correct solution if $h_2(d) = h_1(\frac{2}{3}d_0)$. When $\frac{2}{3}d_0 < d(x) < d_0$, $h_2(d) = -(4/\pi E)d_0\nu F$ is decreasing as the centripetal force $F = F(x)$ is increasing. Consequently, the intersections of $h_1(d)$ with $h_2(d)$ indicate that the diameter of the whirling rod is decreasing because $h_1'(d) = 3d^2 - 2dd_0 > 0$. Hence the intersections between $h_1(d)$ and $h_2(d)$ give the physically correct solution. When $0 < d(x) < \frac{2}{3}d_0$, $h_2(d) = -(4/\pi E)d_0\nu F$ is decreasing as the centripetal force $F = F(x)$ is increasing. Consequently, the intersections of $h_1(d)$ with $h_2(d)$ indicate that the diameter of the whirling rod is increasing because $h_1'(d) = 3d^2 - 2dd_0 < 0$. Hence the intersections between $h_1(d)$ and $h_2(d)$ give the physically incorrect solution. Moreover, if Ω keeps increasing beyond a point such that $h_2(d) < h_1(\frac{2}{3}d_0)$ for some or all $x \in [0, L]$, then $h_1(d)$ and $h_2(d)$ will not intersect each other for those $x \in [0, L]$. Hence no solution will be obtained. As a result, there are no physically correct solutions for Eqs. (8) when $d(x) < \frac{2}{3}d_0$ for some or all $x \in [0, L]$. □

It is at this point, when $d = \frac{2}{3}d_0$, "static instabilities" of the solutions start to set in. For the numerical experiment using the method proposed in Sec. III and the parameters in Sec. IV, numerical solutions become unstable when $d(0)/d_0 < 0.7$, which is slightly different from the $d(0)/d_0 < \frac{2}{3}$ stated in Theorem 2. The difference is probably because of the truncation error of using a finite n in the complete functions approximation of the solution and the numerical round-off error. When $d(0) = 0.7d_0$, the corresponding angular velocity $\Omega \approx 4094$ rad/s from the numerical results compared to 7929 rad/s addressed by Theorem 1 for the linear uniaxial model.

VI. DISCUSSION

It has been noted that the cross-sectional area of a whirling rod in the uniaxial model is assumed to be constant everywhere along the rod at all time. Throughout the derivation of the proposed nonlinear model for a rod in Lagrangian coordinate x, the cross-sectional area is assumed to be circular. Nevertheless, the Poisson's ratio is applicable to noncircular cross-section rods. Recall that Poisson's ratio is the ratio between the lateral strain and the longitudinal strain for uniaxial stress. If each cross-sectional area of a rod under consideration is described by the product of two quantities, namely $l_0(x)$ and $w_0(x)$, then

$$\varepsilon_L(x) = \frac{l(x) - l_0(x)}{l_0(x)} = -\nu\varepsilon(x)$$

and

$$\varepsilon_L(x) = \frac{w(x) - w_0(x)}{w_0(x)} = -\nu\varepsilon(x),$$

where $l(x)$ and $w(x)$ are the new deformed quantities from the original undeformed quantities $l_0(x)$ and $w_0(x)$, respectively. After rearranging the above two equations,

$$l(x) = l_0(x)[1 - \nu\varepsilon(x)]$$

and

$$w(x) = w_0(x)[1 - \nu\varepsilon(x)].$$

As a result,

$$l(x)w(x) = l_0(x)w_0(x)[1 - \nu\varepsilon(x)]^2.$$

Defining the new deformed cross-sectional area $A(x) = l(x)\, w(x)$ and the original undeformed cross-sectional area $A_0(x) = l_0(x)\ w_0(x)$ at location x, then, from the previous equation,

$$A(x) = A_0(x)[1 - \nu\varepsilon(x)]^2.$$

If this expression is embraced in deriving the proposed nonlinear model, Eq. (7) shall be replaced by

$$EA_0(x)[1 - \nu\varepsilon(x)]^2\varepsilon(x) = \int_x^L \rho_0 A_0(x)(x+u)\Omega^2\, dx.$$

Hence the only difference in deriving the proposed nonlinear governing equation for whirling rods with noncircular cross section is to engage the original undeformed cross-sectional area profile $A_0(x)$ on both sides of Eq. (8a).

VII. REMARKS AND CONCLUSIONS

Results for the longitudinal displacements for steadily rotating rods show that the nonlinear strain-displacement relationship should be taken into account at high angular velocity. However, even when made of high-yield steel, which has a yield stress of about 1200 MPa, the corresponding rotational speed according to the linear uniaxial model is 558 rad/s, which is well below the limit 7929 rad/s for the existence and uniqueness of solution described in Sec. V A. The corresponding rotational speed according to the nonlinear model is 551 rad/s, which is well below the limit

4094 rad/s for the existence and uniqueness of solution described in Sec. V B. This indicates that existing materials will fail well before the linear or nonlinear solutions become invalid. Furthermore, beyond the rotational speed at yield stress, Hooke's law ceases to be valid. Hence, for the prediction of longitudinal displacements of steel rods with angular velocities below 558 rad/s, the linear uniaxial model is adequate. Since the linear uniaxial model does not take the variation of cross-sectional area and density of material into account, it is difficult to determine which method to predict the longitudinal natural frequencies is more accurate. Recall that the receptance approach takes the variation of area and density into account for nonlinear deformation. Based on the results obtained in this paper, longitudinal natural frequencies are lowered when the rod is in static rotation compared to its nonrotating stationary state. Rotating rods are in horn shape according to the numerical results. The density of the rotating rod increases towards the free end of the rod. Solutions for the linear uniaxial model with angular velocities rotating up to the yield point of high-strength steel exist and are unique. Numerical solutions for the nonlinear model with angular velocities rotating up to the yield point of high-strength steel exist but are nonunique when $d(x) > 0.7d_0$ for all $x \in [0, L]$. This is because one of the solutions is physically correct and the other solution is physically implausible with the reasoning mentioned in Sec. V B. If the numerical iterations of the nonlinear model converge to physically implausible solutions under the condition $d(x) > 0.7d_0$ for all $x \in [0, L]$, a different n must be attempted in the numerical procedure until a physically correct solution is found. Recall that n is the size of a complete series in Eqs. (10).

[1] C. Venkatesan and V. T. Nagaraj, "On the axial vibrations of rotating bars," J. Sound Vib. **74**, 143–147 (1981).
[2] P. G. Bhuta and J. P. Jones, "On axial vibrations of a whirling bar," J. Acoust. Soc. Am. **35**, 217–221 (1963).
[3] D. H. Hodges and R. R. Bless, "Axial instability of rotating rods revisited," Int. J. Non-Linear Mech. **29**, 879–887 (1994).
[4] E. J. Brunelle, "Stress redistribution and instability of rotating beams and disks," AIAA J. **9**, 758–759 (1971).
[5] G. L. Anderson, "On the extensional and flexural vibrations of rotating bars," Int. J. Non-Linear Mech. **10**, 223–236 (1975).
[6] D. H. Hodges, "On the extensional vibrations of rotating bars," Int. J. Non-Linear Mech. **12**, 293–296 (1977).
[7] D. H. Hodges, "Comments on 'On the axial vibrations of rotating bars'," J. Sound Vib. **87**, 513–515 (1983).
[8] O. M. O'Reilly and J. S. Turcotte, "On the free vibration of a whirling rod," Proceedings of DETC'97, 1997 ASME Design Engineering Technical Conferences, Paper No. DETC97VIB4072, Sacramento, CA, 14–17 September 1997, pp. 1–9.
[9] N. M. Kinkaid, O. M. O'Reilly, and J. S. Turcotte, "On the steady motions of a rotating elastic rod," J. Appl. Mech. **68**, 766–771 (2001).
[10] Y. C. Fung, *A First Course in Continuum Mechanics (Third Edition)* (Prentice Hall, Englewood Cliffs, NJ, 1994).
[11] D. C. Hestermann, R. D. Entwistle, and B. J. Stone, "Axial and torsional receptances for tapered circular shafts, Modal Analysis," Int. J. Anal. Exp. Modal Anal. **11**, 178–193 (1996).
[12] H. Sagan, *Boundary and Eigenvalue Problems in Mathematical Physics* (Dover, New York, 1989), pp. 329–331.

A closed form solution for the dynamic response of finite ribbed plates[a)]

Tian Ran Lin and Jie Pan[b)]
Center for Acoustics, Dynamics, and Vibration, School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009 Australia

(Received 16 March 2005; revised 3 November 2005; accepted 9 November 2005)

A simple and closed form solution for the vibration response of finite ribbed plates to point force/moment excitations is presented in this paper. This solution shows that input mobilities of finite ribbed plates are bounded by the input mobilities of the uncoupled plate and beam that form the ribbed plate. It is found that point force input mobilities of a finite ribbed plate are controlled by the plate bending stiffness when the excitation force is more than a quarter wavelength away from the beam. The input mobilities are mainly dominated by the beam flexural stiffness when the force acts on or very close to the beam, and when the beam flexural stiffness is far greater than the plate bending stiffness. A similar result is found in the moment excitation case when the moment axis is perpendicular to the beam neutral axis (bending moment excitation). In contrast, the input mobilities of the ribbed plate do not vary much from that of the corresponding uncoupled plate when the moment axis parallels to the beam's neutral axis (torsional moment excitation) where the input mobilities are mainly dominated by the plate bending stiffness. The reductions in plate kinetic energy due to beam insertions are discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146111]

PACS number(s): 43.40.Dx, 43.40.Cw [ANN] Pages: 917–925

I. INTRODUCTION

The vibration response of beam stiffened plates has been an interesting research topic for decades. The three landmark papers published in the early 1960s by Maidanik,[1] Lamb,[2] and Ungar[3] constructively highlighted the relevance of the vibration of ribbed plates to important vibroacoustical problems. Maidanik focused on the vibration response of ribbed panels to diffuse sound fields. He found that the ribbing increases the radiation resistance of the panel and therefore encourages the energy exchange between the panel and sound field. His work also significantly contributed to the development of a technique, called Statistical Energy Analysis, which has become a common tool used in the vibroacoustical area. The work by Lamb concentrated on the point force input impedance of an infinite elastic beam attached to a thin elastic plate. The property of the impedance was related to the vibration radiation characteristics from the beam to the plate in terms of the ratio of flexural wave speed on the plate to that on the beam. Strong radiation into plate was found possible when the ratio is less than one, while plate excitation by the beam becomes localized when the ratio is larger than one. His work reveals a strong coupling between the beam and plate at the low frequencies. At very high frequencies, the impedance approaches that of a free beam. Ungar, on the other hand, looked at plate flexural wave transmission through reinforcing beams with specific interests on the dynamic stress concentration induced by wave interference near the beam/plate interfaces. The phenomenon of coincidence transmission of flexural wave in the plate corresponding to its trace matching with flexural and torsional waves in the beam was discovered.

Since Refs. 1–3, many papers on ribbed plate vibration and its interaction with surrounding fluid were published. The effort has been mainly in the areas of developing techniques to solve the coupled beam/plate partial differential equations and using ribbed plates as a model to gain a better understanding of the vibroacoustical properties in complex structures. Techniques, such as the Rayleigh-Ritz energy method,[4–7] finite difference method,[8] transfer matrix method,[9] semimodal decomposition method,[10] and differential quadrature method[11] have been developed to predict the free vibrations of finite ribbed plates. On the other hand, the Fourier transform technique is often employed in solving the vibration problem of infinite ribbed plates with forced excitations.[12–17] In particular, Mead[18] used receptance functions and wave propagation constants in obtaining analytical solutions for a periodic infinite Timonshenko beam and infinite beam stiffened plates. Although the analysis of infinite ribbed plates greatly improved the understanding of the vibration characteristics of ribbed plate structures, the extension of such analysis approaches to finite ribbed plates requires further work by including the effect of boundary reflections.

Through studying the vibration of finite ribbed plates, Mead and Yaman[19] provided an analytical solution for the vibration response of a finite sandwich beam stiffened rectangular plate to line force excitations by the wave approach.

[a)]Some of the results were presented in "Some vibration characteristics of finite ribbed plates," in Proceedings of the 11th International Congress on Sound and Vibration, St. Petersburg, Russia, July 2004.

[b)]Electronic mail: pan@mech.uwa.edu.au

Nevertheless, because the wave coefficients were determined from matrix inversion, the solution was not given in explicit form.

Accompanying the fast advance of computer technologies, the use of finite element analysis (FEA) for the vibration analysis of finite ribbed plate structures has become extensive. Mead *et al.*[20] used the hierarchical finite element model and calculated the natural modes of a flat plate reinforced by grids of beam stiffeners in both plate directions. Effects of boundary conditions on the natural frequencies of stiffened plates and the coincident effect of plate response to the pressure field are discussed. Panali *et al.*[21] used direct finite element formulations for the static and vibration analysis of beam stiffened plates with different boundary conditions. Transverse shear deformations of plates were included in the formulation. Orrenius and Finnveden[22] applied the FEA method in studying wave propagation in ribbed plate structures using a waveguide model. FEA is an effective numerical tool in predicting the vibration characteristics of ribbed plate structures with complex configurations at lower frequencies. However, simple and closed form analytical solutions for the response and input mobility of complex structures (even though they may be approximated solutions) are always sought for a quick and qualitative engineering estimate.

One of the important measures of structure response to an external forcing is the input mobility (impedance). Pinnington and White[23] studied the power transmission from a vibrating machinery to the supporting beam and found that the frequency averaged input mobility of the supporting beam can be approximated by that of the corresponding infinite beam. By extending the work of Lamb,[1] input mobility and power flow of an infinite ribbed plate due to point force and moment excitations applied on the beam were studied by Goyder.[15] He found that for the point force and bending moment excitations at the beam, the beam flexural stiffness initially controls the transmitted power flow into the plate. At distance away from the source location, the power flow in the beam is negligible when compared with the power in the plate. The radiated vibration field from the beam has certain directionality and the maximum vibration response occurs at the coincidence angles $\pm\sin^{-1}(k_b/k_p)$ on the plate where k_b and k_p are, respectively, the flexural wave number of the beam and plate. This result agrees well with those obtained by Cremer *et al.*[24] and Unger[3] using propagation wave approaches. In torsional moment excitation case, it was found that a ribbed plate behaves like a plate at low frequencies or when the beam torsional stiffness and inertia is relatively small. At high frequencies or when the plate bending stiffness is relatively small, the ribbed plate response approaches to the response of a corresponding beam. It was also found that the beam could radiate energy freely to the plate from the driving point at low frequencies.

In this paper, the vibration response of a finite ribbed plate due to point force and moment excitations applied either on the plate or on the beam is analyzed using the well-known modal expansion method. This work is motivated by a practical requirement of estimating input power into the hull structure of a ship due to the excitations of the propeller shafts through the mechanical joints between the shaft and hull. The paper focuses on the derivation of simple and closed form solutions for the input mobilities and kinetic energy of the ribbed plate, so that they might be useful for ship builders for a quick estimate of the input power and structural response. The accuracy of the solutions is evaluated by comparing them with FEA results and limitations of the solutions are also discussed. Then the solutions are used to study the mobility functions with respect to the asymptotic mobilities from the infinite beam and plate systems and to the relative distance between the source and the beam. The general properties of the mobility functions of a finite ribbed plate are summarized.

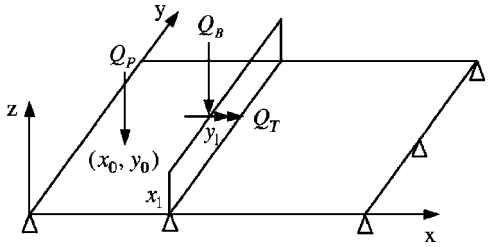


FIG. 1. Model description and the coordinate system.

Given an excitation on the plate, the effect of a stiffening beam on the kinetic energy in the plate is described by the "beam insertion loss."[25] The dependence of the insertion loss on the source location with respect to the beam, and on the frequency is also investigated.

II. GENERAL FORMULATIONS

A ribbed plate model and the associated coordinate system are shown in Fig. 1. In this analysis, we assumed that the stiffened beam is firmly attached to one side of the plate where the beam/plate interface is considered as a nonslip line connection. Simply supported boundary conditions are assumed for all plate edges.

Using the thin plate vibration model, the governing equation of the forced plate bending displacement (W) to the force excitations as shown in Fig. 2(a) can be written as

$$\nabla^4 W - k_p^4 W = \frac{Q_P}{D} - \frac{F}{D}\delta(x - x_1) - \frac{M}{D}\delta'(x - x_1), \tag{1}$$

where D and k_p are the plate bending stiffness and wave number, respectively. Q_p is the external forcing term, F and M are the coupling force and moment at the beam/plate in-

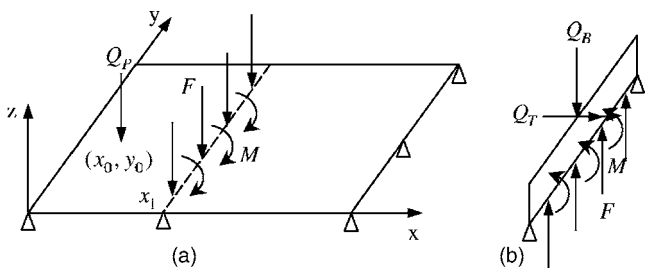


FIG. 2. Illustrations of the plate and beam components of the ribbed plate and the associated external and coupling force excitations for each component.

terface. x_1 is the beam location on the plate and the prime sign ($'$) indicates a spatial derivative.

It is assumed that the beam is symmetrical with respect to its neutral plane so that the flexural and torsional vibration of the beam is uncoupled. The governing equations of beam flexural and torsional displacements (U,θ) to external force (moment) excitations as shown in Fig. 2(b) are

$$\frac{\partial^4 U}{\partial y^4} - k_b^4 U = \frac{Q_B}{B} + \frac{F}{B}, \tag{2}$$

and

$$\frac{\partial^2 \theta}{\partial y^2} + k_t^2 \theta - R\frac{\partial^4 \theta}{\partial y^4} = \frac{Q_T}{T} + \frac{M}{T}, \tag{3}$$

where k_b and k_t are beam flexural and torsional wave numbers, B and T are beam bending and torsional stiffness, respectively. R is the beam warping to torsional stiffness ratio, Q_B and Q_T are the external moments (force) applied on the beam. Equations (1)–(3) can be solved by modal expansions of W, U, and θ as

$$W = \sum_m \sum_n w_{m,n}\phi_m(x)\phi_n(y), \tag{4}$$

$$U = \sum_n u_n \phi_n(y), \tag{5}$$

and

$$\theta = \sum_n \theta_n \phi_n(y), \tag{6}$$

where $\phi_m(x)=\sin(k_m x)$, $\phi_n(y)=\sin(k_n y)$, $k_m=m\pi/L_x$, and $k_n=n\pi/L_y$ are the mode shape functions and modal wave numbers of the simply supported rectangular plate with respect to the two plate edge directions.

Substituting Eq. (4) into Eq. (1) and integrating over the plate surface, one has

$$w_{m,n} = \frac{1}{G_{m,n}}\int_S Q_P \phi_m(x)\phi_n(y)dS - \frac{\phi_m(x_1)}{G_{m,n}}\int_0^{Ly} F\phi_n(y)dy - \frac{\phi_m'(x_1)}{G_{m,n}}\int_0^{Ly} M\phi_n(y)dy, \tag{7}$$

where S is the total plate surface area, $dS=dxdy$, $G_{m,n}=D\Lambda_{m,n}(k_{m,n}^4-k_p^4)$, $\Lambda_{m,n}=L_xL_y/4$, and $k_{m,n}=\sqrt{k_m^2+k_n^2}$ is the modal wave number of the plate.

Similarly, by substituting Eqs. (5) and (6) into Eqs. (2) and (3) and integrating over the beam length, we obtain

$$u_n = \frac{1}{G_n}\left(\int_0^{Ly} Q_B\phi_n(y)dy + \int_0^{Ly} F\phi_n(y)dy\right), \tag{8}$$

and

$$\theta_n = -\frac{1}{G_{Tn}}\left(\int_0^{Ly} Q_T\phi_n(y)dy + \int_0^{Ly} M\phi_n(y)dy\right), \tag{9}$$

where $G_n=B\Lambda_n(k_n^4-k_b^4)$, $G_{Tn}=T\Lambda_n(Rk_n^4+k_n^2-k_t^2)$, $\Lambda_n=L_y/2$.

Applying the two compatibility conditions at the beam/plate interface $[U(y)=W(x_1,y)$ and $\theta(y)=\partial W/\partial x(x_1,y)]$, the modal coupling force and moment at the interface can be determined by

$$\int_0^{Ly} F\phi_n(y)dy = \frac{\alpha_3}{\alpha_1\alpha_3-\alpha_2^2}\left(\sum_m \frac{\phi_m(x_1)\int_S Q_P\phi_m(x)\phi_n(y)dS}{G_{m,n}} - \frac{\int_0^{Ly} Q_B\phi_n(y)dy}{G_n}\right) - \frac{\alpha_2}{\alpha_1\alpha_3-\alpha_2^2}\left(\sum_m \frac{\phi_m'(x_1)\int_S Q_P\phi_m(x)\phi_n(y)dS}{G_{m,n}} + \frac{\int_0^{Ly} Q_T\phi_n(y)dy}{G_{Tn}}\right), \tag{10}$$

$$\int_0^{Ly} M\phi_n(y)dy = -\frac{\alpha_2}{\alpha_1\alpha_3-\alpha_2^2}\left(\sum_m \frac{\phi_m(x_1)\int_S Q_P\phi_m(x)\phi_n(y)dS}{G_{m,n}} - \frac{\int_0^{Ly} Q_B\phi_n(y)dy}{G_n}\right) + \frac{\alpha_1}{\alpha_1\alpha_3-\alpha_2^2}\left(\sum_m \frac{\phi_m'(x_1)\int_S Q_P\phi_m(x)\phi_n(y)dS}{G_{m,n}} + \frac{\int_0^{Ly} Q_T\phi_n(y)dy}{G_{Tn}}\right), \tag{11}$$

where

$$\alpha_1 = \frac{1}{G_n} + \sum_m \frac{\phi_m^2(x_1)}{G_{m,n}}, \quad \alpha_2 = \sum_m \frac{\phi_m(x_1)\phi_m'(x_1)}{G_{m,n}},$$

$$\alpha_3 = -\frac{1}{G_{Tn}} + \sum_m \frac{\phi_m'^2(x_1)}{G_{m,n}}.$$

The modal coefficients for plate (bending) and beam (flexural and torsional) vibrations can now be obtained from Eqs. (7)–(9) using Eqs. (10) and (11). For better illustration, the modal coefficients and plate response to each individual force or moment excitation are given by closed form solutions and discussed in the subsequent analysis.

III. CLOSED FORM SOLUTIONS

A. Ribbed plate response to a point force excitation on the beam

For this excitation, $Q_P=Q_T=0$, $Q_B=F_B\delta(y-y_1)$, where F_B is the external point force applied on the beam at the beam location $y=y_1$. From Eqs. (10) and (11), the modal coupling force and moment at the interface are found to be

$$\int_0^{Ly} F\phi_n(y)dy = -\frac{\alpha_3}{G_n(\alpha_1\alpha_3-\alpha_2^2)}F_B\phi_n(y_1) \tag{12}$$

and

$$\int_0^{Ly} M\phi_n(y)dy = \frac{\alpha_2}{G_n(\alpha_1\alpha_3-\alpha_2^2)}F_B\phi_n(y_1). \tag{13}$$

Equations (12) and (13) are substituted into Eq. (7) to give

$$w_{m,n} = F_B\frac{\phi_n(y_1)[\alpha_3\phi_m(x_1)-\alpha_2\phi_m'(x_1)]}{G_{m,n}G_n(\alpha_1\alpha_3-\alpha_2^2)}. \tag{14}$$

The plate vibration response can now be calculated from Eq. (4).

Further, by applying one of the compatibility conditions at the interface $[U(y)=W(x_1,y)]$, the input mobility of the point force can be obtained from Eqs. (4) and (14) as

$$Y_{in}^{F_B} = \frac{\dot{U}(y_1)}{F_B} = \frac{j\omega}{F_B}\sum_m\sum_n w_{m,n}\phi_m(x_1)\phi_n(y_1), \tag{15}$$

where the dot sign above $U(y_1)$ represents a time derivative.

B. Ribbed plate response to a bending moment excitation on the beam

For this excitation, $Q_P=Q_T=0$, $Q_B=M_B\delta'(y-y_1)$, where M_B is the bending moment applied on the beam. The coupling force and moment at the interface for this excitation case are similar to that given by Eqs. (12) and (13) except that the force term $F_B\phi_n(y_1)$ is now replaced by $M_B\phi_n'(y_1)$. The modal coefficient of the plate response for this excitation case becomes

$$w_{m,n} = M_B\frac{\phi_n'(y_1)[\alpha_3\phi_m(x_1)-\alpha_2\phi_m'(x_1)]}{G_{m,n}G_n(\alpha_1\alpha_3-\alpha_2^2)}. \tag{16}$$

As a result, the bending moment input mobility of the ribbed plate is

$$Y_x^{M_B} = \frac{\frac{\partial\dot{U}}{\partial y}(y_1)}{M_B} = \frac{j\omega}{M_B}\sum_m\sum_n w_{m,n}\phi_m(x_1)\phi_n'(y_1). \tag{17}$$

C. Ribbed plate response to a torsional moment excitation on the beam

For this excitation, $Q_P=Q_B=0$, $Q_T=M_T\delta(y-y_1)$, where M_T is the external torsional moment applied on the beam. The modal coefficient of the plate response is given by

$$w_{m,n} = M_T\frac{\phi_n(y_1)[\alpha_2\phi_m(x_1)-\alpha_1\phi_m'(x_1)]}{G_{m,n}G_{Tn}(\alpha_1\alpha_3-\alpha_2^2)}. \tag{18}$$

Applying the other compatibility condition at the interface $[\theta(y)=(\partial W/\partial x)(x_1,y)]$, the input mobility of the torsional moment is obtained as

$$Y_{in}^{M_T} = \frac{\dot{\theta}(y_1)}{M_T} = \frac{j\omega}{M_T}\sum_m\sum_n w_{m,n}\phi_m'(x_1)\phi_n(y_1). \tag{19}$$

D. Ribbed plate response to a point force excitation on the plate

For this excitation case, $Q_B=Q_T=0$, $Q_p=F_P\delta(x-x_0)\delta(y-y_0)$, where F_P is the external point force applied on the plate. The modal coefficient of the plate response is obtained as

$$w_{m,n} = F_P\frac{\phi_n(y_0)}{G_{m,n}}\left[\phi_m(x_0) - \frac{\alpha_3\alpha_4-\alpha_2\alpha_5}{\alpha_1\alpha_3-\alpha_2^2}\phi_m(x_1) - \frac{\alpha_1\alpha_5-\alpha_2\alpha_4}{\alpha_1\alpha_3-\alpha_2^2}\phi_m'(x_1)\right],$$

where

$$\alpha_4 = \sum_m \frac{\phi_m(x_0)\phi_m(x_1)}{G_{m,n}} \tag{20}$$

and

$$\alpha_5 = \sum_m \frac{\phi_m(x_0)\phi_m'(x_1)}{G_{m,n}}.$$

The first term in the square bracket of Eq. (20) is the modal coefficient of the corresponding uncoupled plate to the same excitation and the last two terms are the contributions to the plate response by the force and moment coupling of the stiffened beam. It should be noted that Eq. (20) becomes Eq. (14) when $x_0=x_1$, $y_0=y_1$. Therefore, the excitation case where a point force is applied on the beam is simply a special case of this excitation.

The input mobility of the ribbed plate for this case is given by

$$Y_{in}^{F_P} = \frac{j\omega}{F_P} \sum_m \sum_n w_{m,n} \phi_m(x_0) \phi_n(y_0). \tag{21}$$

E. Ribbed plate response to a moment excitation on the plate

Letting $Q_T = Q_B = 0$, $Q_P = -M_0 \cos\varphi \delta(x-x_0)\delta'(y-y_0) + M_0 \sin\varphi \delta'(x-x_0)\delta(y-y_0)$,[26] where M_0 is the amplitude of the external moment and φ is the angle of the moment axis with respect to the x-axis of the coordinate system, and applying the compatibility conditions at the interface, the modal coefficients of the plate response is obtained as

$$\begin{aligned} w_{m,n} = -\frac{M_0}{G_{m,n}} \Bigg\{ & \cos\varphi \phi_n'(y_0) \Bigg[\phi_m(x_0) \\ & - \frac{\alpha_3\alpha_4 - \alpha_2\alpha_5}{\alpha_1\alpha_3 - \alpha_2^2} \phi_m(x_1) - \frac{\alpha_1\alpha_5 - \alpha_2\alpha_4}{\alpha_1\alpha_3 - \alpha_2^2} \phi_m'(x_1) \Bigg] \\ & - \sin\varphi \phi_n(y_0) \Bigg[\phi_m'(x_0) - \frac{\alpha_3\alpha_6 - \alpha_2\alpha_7}{\alpha_1\alpha_3 - \alpha_2^2} \phi_m(x_1) \\ & - \frac{\alpha_1\alpha_7 - \alpha_2\alpha_6}{\alpha_1\alpha_3 - \alpha_2^2} \phi_m'(x_1) \Bigg] \Bigg\}, \end{aligned} \tag{22}$$

where

$$\alpha_6 = \sum_m \frac{\phi_m'(x_0)\phi_m(x_1)}{G_{m,n}}, \quad \alpha_7 = \sum_m \frac{\phi_m'(x_0)\phi_m'(x_1)}{G_{m,n}}.$$

There are two input mobility components for this excitation case; each associates with one of the two angular velocity components at the source location $[\partial\dot{W}/\partial x(x_0,y_0)$ and $\partial\dot{W}/\partial y(x_0,y_0)]$ and are given, respectively, by

$$Y_x^{M_0} = \frac{\frac{\partial\dot{W}}{\partial y}(x_0,y_0)}{M_0} = \frac{j\omega}{M_0} \sum_m \sum_n w_{m,n} \phi_m(x_0) \phi_n'(y_0) \tag{23}$$

and

$$Y_y^{M_0} = \frac{\frac{\partial\dot{W}}{\partial x}(x_0,y_0)}{M_0} = \frac{j\omega}{M_0} \sum_m \sum_n w_{m,n} \phi_m'(x_0) \phi_n(y_0). \tag{24}$$

F. Vibration energy distributions

For all excitation cases, the time averaged, steady state kinetic energy of plate flexural vibration is calculated by

$$\langle T \rangle_P = \frac{1}{2} \int_S \rho_s \dot{W} \dot{W}^* dS = \frac{\rho_s \Lambda_{m,n} \omega^2}{2} \sum_m \sum_n |w_{m,n}|^2, \tag{25}$$

where ρ_s is the plate surface mass and the asterisk $(*)$ indicates a complex conjugate.

Similarly, by utilizing the compatibility conditions at the interface, the time averaged, steady state energy distributions of beam flexural and torsional vibrations can be calculated as

$$\langle T \rangle_b = \frac{1}{2} \int_0^{Ly} \rho_L \dot{U} \dot{U}^* dy = \frac{\rho_L \Lambda_n \omega^2}{2} \sum_n \left| \sum_m w_{m,n} \phi_m(x_1) \right|^2, \tag{26}$$

is the beam flexural vibration energy, and

$$\begin{aligned} \langle T \rangle_t &= \frac{1}{2} \int_0^{Ly} \rho_b I_P \dot{\theta} \dot{\theta}^* dy \\ &= \frac{\rho_b I_P \Lambda_n \omega^2}{2} \sum_n \left| \sum_m w_{m,n} \phi_m'(x_1) \right|^2, \end{aligned} \tag{27}$$

is the beam torsional vibration energy, where ρ_L is the beam mass per unit length and $\rho_b I_p$ is the mass moment of inertia per unit length of the beam.

IV. NUMERICAL SIMULATION AND RESULT DISCUSSIONS

In the numerical simulation, it is assumed that the plate and the beam are both made of aluminum, with material properties $E = 7.1 \times 10^{10}$ N/m^2, $\rho = 2660$ kg/m^3, $\nu = 0.3$. The plate has a surface area of 3.6×1 m^2 and thickness of 8 mm, while the beam is 1 m long with a flat rectangular cross section of $A = 50$ mm $\times 5$ mm. The size and the material properties of the plate and the beam are chosen based on one of the ship hull plates of a 30 m crew vessel currently under investigation. The ship hull plate is bounded by two parallel frames (1 m apart), the keel and a perpendicular ship hull plate so that simply supported boundary conditions can be assumed for all four plate edges. It is further assumed that the plate has a constant internal loss factor $\eta_p = 0.03$ while the beam has $\eta_b = 0.01$. For all calculations in the subsequent analysis, the source locations are fixed at $(x_0, y_0) = (1.4\text{ m}, 0.32\text{ m})$. Force excitation applied on the beam is achieved by moving the beam position to the source location by letting $x_1 = 1.4$ m in the simulation.

The vibration response of the ribbed plate is calculated by including 1000 uncoupled plate modes and 20 uncoupled beam modes in the modal summation. It is also calculated separately by finite element analysis (FEA) using the commercial software, MSC/NASTRAN. 2250 CQUAD4 thin shell elements and 25 CBAR beam elements are used in the FEA model where the modal frequency response analysis module provided by MSC/NASTRAN is adopted by including 500 structural modes in the analysis. The typical structural characteristic function of input mobility is chosen to represent the ribbed plate response to external force (moment) excitations in this study where only the real part of the input mobility is shown in all figures.

A. Prediction accuracy

The accuracy of the proposed method is examined first by comparing the predicted input mobilities of the ribbed plate to that of FEA simulations for the three excitation cases: (a) point force excitation on the beam; (b) torsional moment excitation on the beam; and (c) point force excitation on the plate. The beam is located at the plate location $x_1 = 1.4$ m for the first two excitation cases and is located at

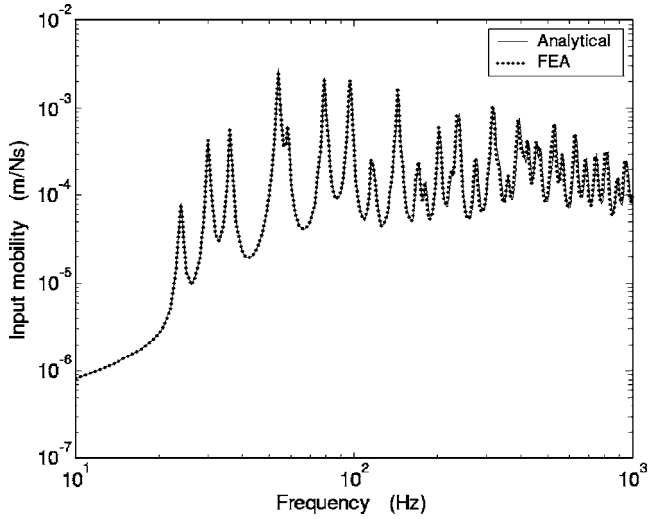


FIG. 3. Input mobility of the ribbed plate to point force excitation on the beam. (—) modal expansion method, (···) FEA.

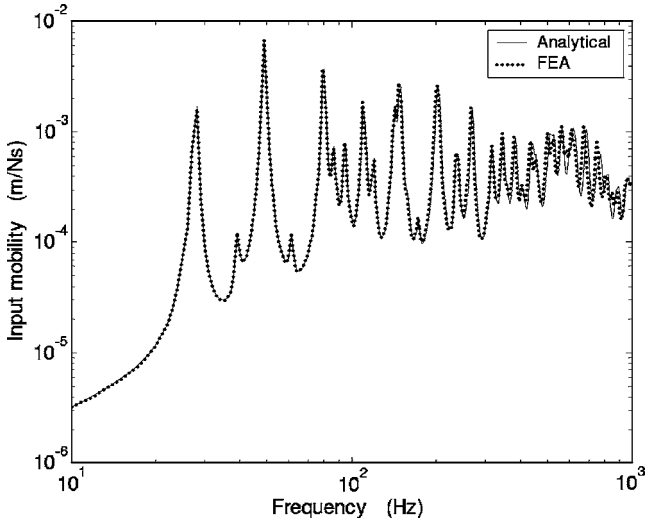


FIG. 5. Input mobility of the ribbed plate to point force excitation on the plate. (—) modal expansion method, (···) FEA.

$x_1 = 1.6$ m for the third excitation case in the simulation. The results are shown in Figs. 3–5, respectively. Good agreement is found in the whole frequency range of investigation for all three excitation cases.

It has to be noted that the method developed here is limited by the assumption made earlier in the analysis where the beam/plate interface is regarded as single line connection. It is found in the simulation that such an assumption is only valid when the beam width is not greater than the plate thickness, which is adopted in many designs of the ship hull plates. For plates reinforced by beams whose beam width is greater than the plate thickness, the current method underestimates the plate impedance to the beam torsional vibration, which leads to the over prediction of the ribbed plate response. However, an approximate solution for such ribbed plates can still be obtained by the modal expansion solution if the source is not a torsional moment and is applied on the beam or on the plate at more than a quarter wavelength away from the beam. For such excitations, it is found that the contribution of beam torsional vibration to ribbed plate response is negligibly small, and can be ignored in the analysis. Therefore, only shear force at the interface needs to be considered, and the ribbed plate response can once again be solved by the modal expansion solution using Euler-Bernoulli beam formulation for the stiffened beam.

B. Characteristics of the ribbed plate input mobility

Input mobilities of the ribbed plate are studied by moving the stiffened beam from the initial location at $x_1 = 1.4$ m to the two plate locations at $x_1 = 1.45$ m and $x_1 = 1.6$ m for the two excitations, point force and moment excitations.

1. Point force input mobility

Figure 6 shows the point force input mobilities of the ribbed plate together with that of the corresponding uncoupled plate and beam of finite and infinite sizes. It is shown that the input mobilities of the ribbed plate are bounded by those of the corresponding uncoupled plate and

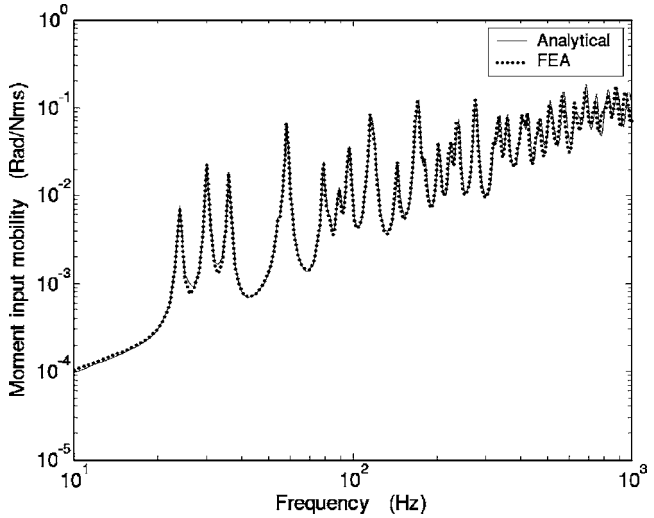


FIG. 4. Input mobility of the ribbed plate to torsional moment excitation on the beam. (—) modal expansion method, (···) FEA.

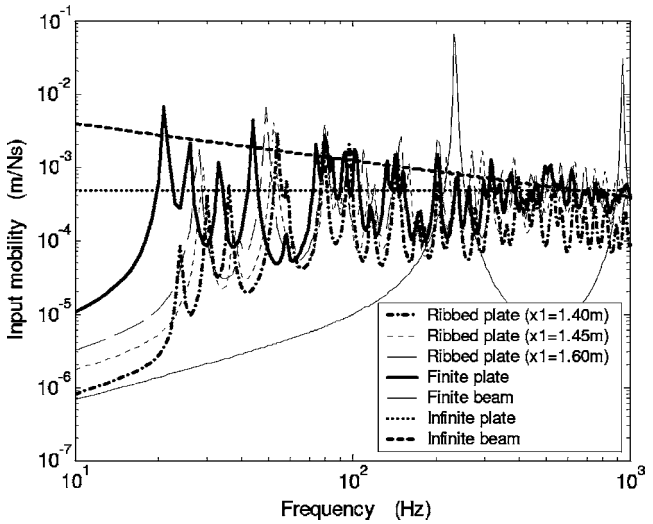


FIG. 6. Input mobilities of the ribbed plate to point force excitations.

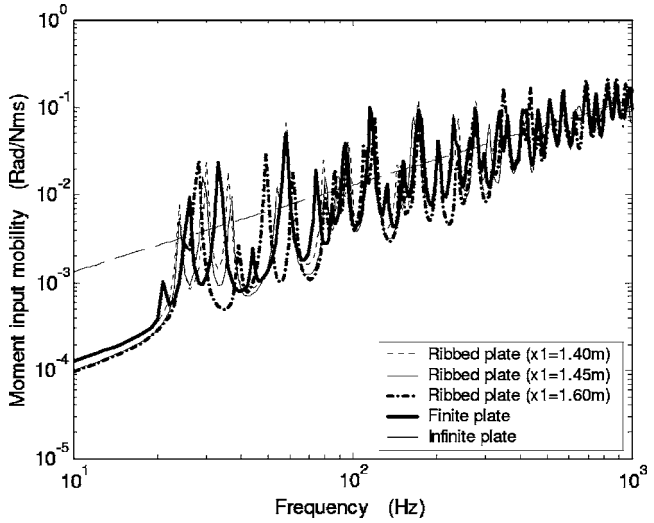


FIG. 7. Input mobilities of the ribbed plate to torsional moment excitations.

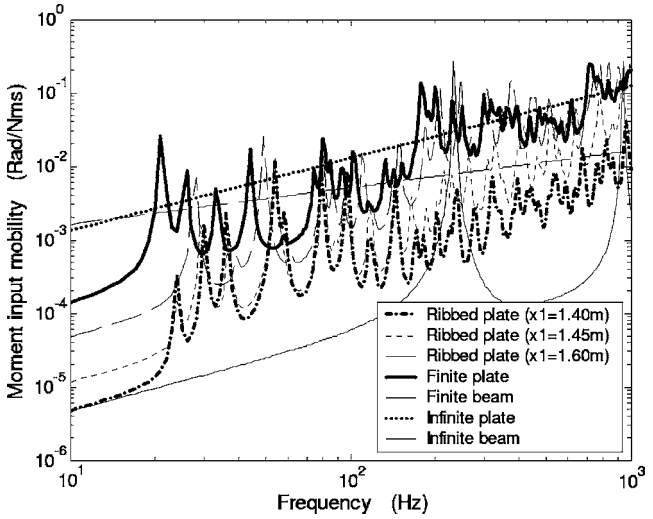


FIG. 8. Input mobilities of the ribbed plate to bending moment excitations.

beam in general. The plate bending stiffness controls the input mobility when the beam is more than a quarter wavelength away from the point force where the input mobility of the ribbed plate is similar to that of the corresponding uncoupled plate. The input mobility of the corresponding infinite plate represents the upper bound of the frequency averaged input mobility of the ribbed plate. Nevertheless, due to the relatively smaller beam/plate flexural stiffness ratio ($B/D \approx 5$) of the ribbed plate, the stiffness term of the input mobility is not only controlled by the beam flexural stiffness but also greatly influenced by the plate bending stiffness when the point force acts on the beam. Therefore, the lower bound of the frequency averaged input mobility of the ribbed plate can not be represented by that of the corresponding infinite beam for this case. It is noted that the observations made here are independent of the y-coordinate of the source when the source is located at more than a quarter wavelength away from the plate edges in this direction.

2. Moment input mobility

The input mobilities of the ribbed plate to bending and torsional moment excitations are considered separately where only the principle component of the moment input mobility (the component that shares the same axis with the external moment either at $\varphi = 0°$ or at $\varphi = 90°$) is examined.

Figure 7 shows the torsional moment input mobilities of the ribbed plate together with that of the corresponding uncoupled plate of finite and infinite extents. It is found that the input mobility of the ribbed plate for this excitation case does not vary much from that of the corresponding uncoupled plate regardless of the location of the stiffened beam. This is because the stiffness term of input mobility for this excitation case is dominated by the plate bending stiffness, which is much greater than the beam torsional stiffness ($D/T > 60$). It is illustrated that the frequency averaged input mobilities for all the three beam/plate configurations increase with increasing frequency and they can be well represented by that of the corresponding infinite plate.

Similar to what is found in point force excitation cases, the bending moment input mobilities of the ribbed plate are also bounded by that of the corresponding uncoupled plate and beam, which is shown in Fig. 8. The frequency averaged input mobilities increase with increasing frequency and can be represented by that of the corresponding infinite plate when the beam is more than a quarter wavelength away from the source.

C. Beam insertion loss to plate response

The reduction of the plate vibration response due to the insertion of a stiffened beam may be measured by using the concept of insertion loss as defined in room acoustics,[27] which is given as

$$\mathrm{IL} = 10 \log_{10} \frac{\langle T \rangle_P^0}{\langle T \rangle_P}, \tag{28}$$

where $\langle T \rangle_P$ is the plate kinetic energy of the ribbed plate given by Eq. (25) and $\langle T \rangle_P^0$ is the kinetic energy of the corresponding uncoupled plate for the same excitation.

The insertion losses corresponding to the three beam locations for cases of point force excitation on the plate are shown in Fig. 9. It is found that large vibration energy reductions can be achieved at low frequencies when the beam is inserted directly between the source and the plate or very close to the source. However, negative reduction is possible even at low frequencies due to the small beam/plate flexural stiffness ratio and the shifted resonant frequencies of the structure after beam insertion. When the stiffened beam is attached to the plate away from the source, the vibration reduction will depend on the distance between the beam and the source. By moving the beam away from the source, the overall insertion loss decreases due to the decreased influence of beam stiffness to the input mobility. The band averaged vibration reduction at high frequencies tends to zero when the beam is more than a quarter wavelength away from the source (i.e., at frequencies greater than 120 Hz for the beam location at $x_1 = 1.6$ m).

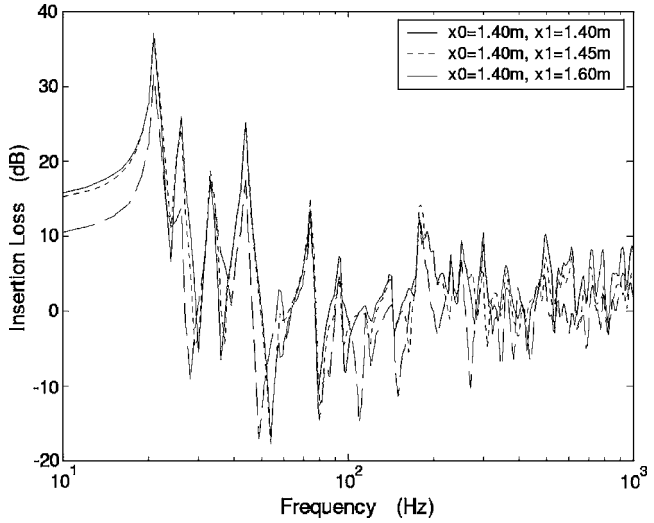


FIG. 9. Beam insertion losses to plate vibration energy due to point force excitation on the plate (and on the beam).

The beam insertion losses in the bending moment excitation case are similar to those of the point force excitation case, and are not shown here. On the other hand, Fig. 10 shows the beam insertion losses in the torsional moment excitation case. It is found that vibration reduction at frequencies below the first system resonant frequency after beam insertion is eventually zero because of the large plate bending to beam torsional stiffness ratio. The reduction also tends to zero at high frequencies. In between, general reductions can be observed at frequencies corresponding to the uncoupled plate natural frequencies while negative reductions are found at frequencies corresponding to the shifted peak frequencies after beam insertion. In general, the insertion of the stiffened beam on the plate in torsional excitation cases only changes the peak response frequencies but not the overall frequency averaged plate vibration energy.

V. CONCLUSIONS

A modal expansion method technique is employed in this paper to obtain an analytical solution for the vibration response of a finite ribbed plate to point force and moment excitations where closed form solutions of the response are obtained in terms of the input mobilities and kinetic energy of the ribbed plate. It is found that the point force and bending moment input mobilities of the ribbed plate are bounded by those of the corresponding uncoupled plate and beam that form the ribbed plate system. They are controlled by the plate bending stiffness when the source is more than a quarter wavelength away from the beam and are mainly dominated by the beam flexural stiffness when the source acts on the beam or very closed to the beam. It is also shown that the torsional moment input mobility of the ribbed plate does not vary much from that of the corresponding uncoupled plate.

Large plate vibration reductions can be found at low frequencies after the insertion of the stiffened beam on the plate in the cases of point force and bending moment excitations. Negative reduction can be observed in the frequency bands associated with some of the shifted peak frequencies after beam insertion. The vibration reduction tends to zero at high frequencies when the beam is more than a quarter of wavelength away from the source. In contrast, the insertion of the beam in torsional excitation case only changes the peak response frequencies but not the overall frequency averaged plate vibration energy.

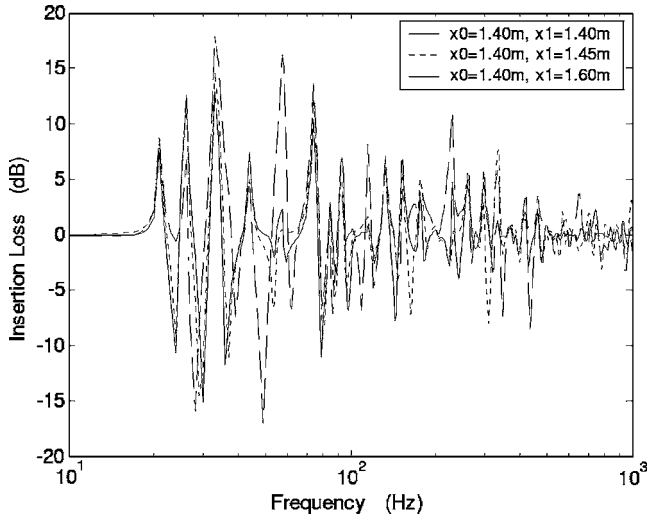


FIG. 10. Beam insertion losses to plate vibration energy due to torsional moment excitations on the plate (and on the beam).

ACKNOWLEDGMENTS

The financial support from the Australian Research Council and Strategic Marine Pty. Ltd. for this work is gratefully acknowledged.

[1]G. Maidanik, "Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **34**, 809–826 (1962).
[2]G. L. Lamb, "Input impedance of a beam coupled to a plate," J. Acoust. Soc. Am. **33**, 628–633 (1961).
[3]E. E. Ungar, "Transmission of plate flexural waves through reinforcing beams; dynamic stress concentration," J. Acoust. Soc. Am. **33**, 633–639 (1961).
[4]C. L. Kirk, "Vibration characteristics of stiffened plates," J. Mech. Eng. Sci. **2**, 242–253 (1960).
[5]D. J. Mead and S. Parthan, "Free wave propagation in two dimensional periodic plates," J. Sound Vib. **64**, 325–348 (1979).
[6]P. A. A. Laura and R. H. Gutierrez, "A note on transverse vibrations of stiffened rectangular plates with edges elastically restrained against rotation," J. Sound Vib. **78**, 139–144 (1981).
[7]J. R. Wu and W. H. Liu, "Vibration of rectangular plates with edge restraints and intermediate stiffeners," J. Sound Vib. **123**, 103–113 (1988).
[8]T. Wah, "Vibration of Stiffened Plates," Aeronaut. Q. **15**, 185–198 (1964).
[9]C. A. Mercer and Miss C. Seavey, "Prediction of natural frequencies and normal modes of skin-stringer panel rows," J. Sound Vib. **6**, 149–162 (1967).
[10]B. R. Long, "A stiffness-type analysis of the vibration of a class of stiffened plates," J. Sound Vib. **16**, 323–335 (1971).
[11]H. Zeng and C. W. Bert, "A differential quadrature analysis of vibration for rectangular stiffened plates," J. Sound Vib. **241**, 247–252 (2001).
[12]S. I. Kovinskaya and A. S. Nikiforov, "Flexural wave fields in infinite beam-reinforced plates under point excitation," Sov. Phys. Acoust. **19**, 32–35 (1973).
[13]M. L. Rumerman, "Vibration and wave propagation in ribbed plates," J. Acoust. Soc. Am. **57**, 370–373 (1975).
[14]G. F. Lin and S. I. Hayek, "Acoustic radiation from point excited rib-reinforced plate," J. Acoust. Soc. Am. **62**, 72–83 (1977).
[15]H. G. D. Goyder, "Vibration analysis using experimental data and approximate methods with consideration of power flow from machinery into built-up structures," Ph.D. thesis, University of Southampton, 1978.
[16]B. R. Mace, "Periodically stiffened fluid loaded plates. II. Response to line and point forces," J. Sound Vib. **73**, 487–504 (1980).

[17]G. P. Eatwell and D. Butler, "The response of a fluid-loaded beam-stiffened plate," J. Sound Vib. **84**, 371–388 (1982).
[18]D. J. Mead, "A new method of analyzing wave propagation in periodic structures: Applications to periodic Timoshenko beams and stiffened plates," J. Sound Vib. **104**, 9–27 (1986).
[19]D. J. Mead and Y. Yaman, "The harmonic response of rectangular sandwich plates with multiple stiffening: A flexural wave analysis," J. Sound Vib. **145**, 409–428 (1991).
[20]D. J. Mead, D. C. Zhu, and N. S. Bardell, "Free vibration of an orthogonally stiffened flat plate," J. Sound Vib. **127**, 19–48 (1988).
[21]G. S. Palani, N. R. Iyer, and T. V. S. R. Appa Rao, "An efficient finite element model for static and vibration analysis of plates with arbitrary located eccentric stiffeners," J. Sound Vib. **166**, 409–427 (1993).
[22]U. Orrenius and S. Finnveden, "Calculation of wave propagation in rib-stiffened plate structures," J. Sound Vib. **198**, 203–224 (1996).
[23]R. J. Pinnington and R. G. White, "Power flow through machine isolators to resonant and nonresonant beams," J. Sound Vib. **75**, 179–197 (1981).
[24]L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound* (Springer-Verlag, Berlin, 1988), Chap. 5, pp. 442–450.
[25]T. R. Lin and J. Pan, "Some vibration characteristics of finite ribbed plates," in Proceeding of the 11th International Congress on Sound and Vibration, St. Petersburg, Russia, 2004, pp. 3369–3376.
[26]J. Pan, J. Q. Pan, and C. H. Hansen, "Total power flow from a vibrating rigid body to a thin panel through multiple elastic mounts," J. Acoust. Soc. Am. **92**, 895–907 (1992).
[27]M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers* (Cambridge University Press, Cambridge 1989), Chap. 4, pp. 288, 293.

Hybrid vibration isolator: Single axis control study

Peter C. Herdic,[a] Robert D. Corsaro, Brian H. Houston, and Robert M. Baden
Physical Acoustics Branch, Naval Research Laboratory, Washington, DC 20375-5320

(Received 19 October 2005; accepted 16 November 2005)

Results are presented for a laboratory study of a compact, single-axis hybrid (active-passive) vibration isolator. The passive system component demonstrates a very high level of vibration isolation at frequencies roughly a factor of 3 above the fundamental system resonance. The active component complements the system by significantly reducing the transmitted vibration levels at lower frequencies, where the passive-only system is ineffective. The device consists of three basic components, a passive compliant spring, force and velocity sensing, and a piezoelectric actuation layer. The experimental system is typically excited at 50 Hz with response characteristics measured over the band from ~10 to 2000 Hz. The isolation performance is evaluated for an optimized passive stage as well as for all relevant hybrid layer configurations. The optimal physical control law is determined by identifying positions in the device stack where actuation and sensing are most effective at minimizing the downstream base velocity and power flow through the mount. Local force and velocity minimization are implemented via a least mean squared adaptive control filter. The role of harmonic distortion and actuator nonlinearity is examined by comparing the system performance and out-of-band enhancement obtained using PZT-4, PZT-5H, and single-crystal PMN-PT actuator materials. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2149843]

PACS number(s): 43.40.Tm [JGM] Pages: 926–936

I. INTRODUCTION

Hybrid systems contain both passive and active components which work together sharing the task of isolating vibrations generated by vibrating machinery. This synergy offers considerable potential and advantage by promising higher levels of performance than are practical with passive-only or active-only systems.

The passive component is typically a conventional, compliant rubber spring. When loaded by a mass, this system behaves as a simple spring-mass system with a fundamental resonance frequency. Den Hartog,[1] Ungar[2] and others describe such systems that provide systematic improvement in isolation at frequencies progressively higher than this resonance. For machinery vibrations in the 50 Hz range, such isolators typically provide moderate levels (10–15 dB) of vibration isolation.[3] Higher levels require that the system resonance be much lower than the operating frequency of interest. At these very low frequencies, the requirement of a very low stiffness spring leads a practical issue associated with supporting the machinery in a stable manner.

Active vibration isolation control does not suffer from this issue since typically all components are stiff.[4] In these systems, an actuator supplies a displacement or force that attempts to decouple the vibration source from the platform. Sensors are used to monitor the source and error residuals, and the control system uses this sensor information to determine the optimum drive signal to deliver to the control actuator. Typically, adaptive least mean squared (LMS) control algorithms[5] are used to minimize the error sensor for the required control law. Such systems are also capable of reliably and robustly providing moderate levels (15–30 dB) of isolation.[3,6] However when configured for higher performance levels, such systems often have robustness issues. These issues are related to the susceptibility of the control system to instabilities, the presence of nonlinearities, and the possibility of out-of-band enhancement.

The hybrid approach combines active and passive components into a single device to provide higher levels of isolation.[3,6–8] At lower frequencies, the passive element provides only a low, baseline level of isolation to the system. When activated, the active isolation system attempts to completely control the dynamics of the system, yielding reduction of transmitted vibration levels beyond that of the passive system alone. As found in this study, at frequencies above approximately three times the system fundamental, the passive isolation stage is very effective and dominates the total level of performance. The passive element also plays an important role as buffer between the active element and the supporting base structure. This buffer will treat residual levels left over from the active system, and in the event of active system shutdown, this system will revert to a safe passive-only isolation system.

This investigation presents results of a laboratory study to evaluate the level of vibration isolation performance provided by a hybrid isolator. Unlike previous studies, it provides a detailed examination of the component arrangement (sensors, actuator, and passive element) to identify an optimal physical control law. It also includes studies exploring the influence of actuator linearity on system control, identifying actuator materials that inherently reduce the associated level of harmonic distortion and out-of-band enhancement.

The results presented here are restricted to the case of a single hybrid isolator. In a practical installation, three or four such hybrid isolators would be required to adequately support the vibrating machinery. The approach considered here

[a]Also with SFA Inc., Largo, MD 20774.

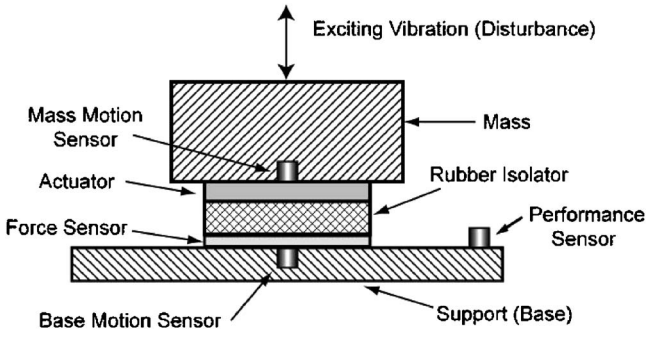


FIG. 1. Conceptual illustration of the hybrid system components.

addresses each isolator as an independent system. This is commonly referred to as local (rather than global) control where separate mounts do not have interconnected controllers. Local control is possible because the measured cross-coupling levels between mounts are negligible for the system in this study. This local approach is advantageous in practical applications since it greatly simplifies the control system. It has a further advantage in this study, in that, essentially all of the information needed can be obtained by studying the dynamics and controllability of just a single unit.

It is also important to note that vibration isolation is truly a three-dimensional problem where a three-axis device is required. That is, the two in-plane components need to be addressed as well as the normal component. For example, in Navy applications, often the in-plane components of the mount are normal to the ship's hull which can lead to high noise radiation levels. In this paper, results of the single-axis control experiments are presented along with the guidance they provided for studies of a three-axis hybrid isolator, which will be the subject of a future paper.

II. HYBRID DEVICE DESIGN

The components of the hybrid isolator are illustrated in Fig. 1. A design goal is to make the system compact enough to fit within a small amount of vertical space, for example, as shelf supports of an electronics cabinet. To achieve this goal, the design uses a conventional passive rubber isolation layer in series with an actuator. The series design requires an actuator with low displacement stroke and high force at low frequencies. A piezoelectric actuator achieves this goal in a small compact package. The device also includes sensors to monitor the force (at least at one location in the path) and the motion of both the driven mass and the base support. Using information provided by these sensors, the control signal is generated to drive the actuator in such a manner as to reduce the coupling between the disturbance source and the base support.

The device sensors and actuators are essentially in-line and colocated. In this study the force sensor and actuator are fabricated from piezoelectric materials. The motion (i.e., velocity or displacement) sensor is a commercial low noise accelerometer. The control is implemented in this study to minimize either velocity or force; however, the approach could be extended to use combinations of the sensor outputs to study impedance or power flow control.

The passive rubber isolator serves four purposes, (1) it provides high isolation at mid to high frequencies, (2) in practical applications, it provides a safety backup or failsafe feature should the active component fail or become unstable, (3) it assigns the spring constant of the system, forcing a predictable and well-behaved relationship between velocity and displacement, and (4) it provides lateral decoupling, which is not otherwise addressed in this single-axis device. The characteristics of the passive rubber isolator were selected to approximate those of one commonly used in our laboratory. The specifications of this isolator are that it supports a load mass of 5 to 6 kg and has a dynamic spring constant of 37 000 N/m. When mounted on a rigid support, this passive spring-mass system provides a system resonance near 13 Hz.

For the active system, the control frequency band is from 25 to 500 Hz. Emphasis is placed on controlling narrow-band machinery noise, which typically occurs near 50 Hz. The maximum disturbance force of interest in this study is ±1100 dynes, corresponding to that produced by equipment in one of our electronic enclosures. Using the spring constant of the passive component, this corresponds to a relatively small displacement of only ±0.30 microns.

The actuator selected must be capable of controlling this motion. The force-displacement-voltage relations for the actuators considered here are approximately linear and hence can be described using two common parameters; the piezoelectric charge constant d_{33} of the material, and the blocked force F_b which is a function of material modulus and actuator dimensions. A resulting expression for the axial displacement delivered by a piezoelectric actuator is

$$\Delta z \equiv \left(1 - \frac{F}{F_b}\right) d_{33} \Delta V, \tag{1}$$

where under low load (F) conditions, the axial displacement Δz is just the product of the piezoelectric charge constant and the applied voltage. Operation in this simple region of the force-displacement graph requires that the maximum force F delivered by the actuator (1100 dynes) be only a small fraction (i.e., 1%) of the blocked force capability of the actuator. The actuator in our application is essentially mounted between a small mass of 5.6 kg and against the soft backing of the compliant passive element, and therefore, it is indeed much lower than the blocked force.

The maximum drive voltage available for this study was chosen as ±300 V for safety considerations in a practical application. Hence when driven at this voltage, the actuator must deliver at least ±0.30 microns displacement. The maximum acceleration produced by this displacement is 0.30 g, which occurs at the highest frequency in this operating band. The maximum acceleration at the test frequency of 50 Hz is only 0.003 g.

A. Actuator

The selected type of actuator design is a simple piezoelectric stack configuration. Piezoceramics are considered good choices for actuation elements when high force and only modest displacement is required. To achieve the re-

quired (0.30 micron) axial displacement using only 300 V, the effective piezoelectric constant of the stack must be at least 1000 pm/V. While piezoelectric actuators are typically considered linear devices, their small inherent nonlinearity presents a complication in active control systems. This is clearly seen when operating the system at a single frequency, where any intrinsic or extrinsic [Eq. (1)] nonlinearity will generate harmonic distortion that can lead to out-of-band enhancements.

Commonly available lead zirconate titanate (PZT) piezoelectric materials fall into three general classes, called "hard," "soft," and "single crystal." Using the common Morgan-Matroc[9] "PZT" designations, typical "hard" PZT formulations are PZT-4 or PZT-8. Typical "soft" formulations are PZT-5A or -5H. Single-crystal materials are relatively new, developed largely under recent ONR and DARPA programs.

As indicated in Eq. (1), the maximum displacement available will depend on the maximum drive voltage. This drive voltage must not present an electric field strength that approaches the depolarization limit of the material, or partial depolarization may occur. The limiting values can be found in manufacturer's literature[9,10] where "hard" PZT materials typically have ac depolarization limits near 600 V/mm, and "soft" materials are typically near 150 V/mm. Single crystal materials typically have positive field limits similar to those of the "soft" materials, but have lower negative coercive field limits; hence they are often used with a positive bias voltage.

Damping (dielectric loss, or tan δ) provides a measure of loss and self-heating. Its value is a function of the electric field strength. Useful relations are given by Eyster[11] for two PZT types based on a best fit to the data. In MKS units these relations become

$$\text{PZT-5H}, \quad \tan\delta = 0.0165 + 0.508E + 0.774E^2, \tag{2}$$

$$\text{PZT-8}, \quad \tan\delta = 0.00585 + 0.0065E + 0.0394E^2, \tag{3}$$

where the field strength E is in units of volts/micron. Values from Eyster's equations can be compared with those appearing in the various manufacturers' literature (i.e., Edo Corp., Sensor Technologies Limited, Channel Industries). It is observed that values calculated for PZT-8 show satisfactory agreement with the manufacturer's listed values, and also give values that reasonably approximate those listed for other "hard" materials, such as PZT-4. Similarly those calculated for PZT-5H also approximate values listed in the manufacturer's literature for other "soft" materials such as PZT-5A.

Linearity is defined in various ways, and consequently values provided by different studies often do not agree. However regardless of the method, materials with lower damping are generally more linear. Hence the soft PZT materials have nonlinearity values that are much larger than those of the "hard" PZT materials. This will be further discussed later in this paper.

The actuator used in the principle portion of this study was fabricated using the generally more linear "hard" material, a PZT-4 type (Channel Industries, type 5400). This particular choice of hard material was partially influenced by wafer availability. The piezoelectric charge constant of this material is approximately 285 pm/V. The stack we fabricated used 0.508 mm thick wafers. For this wafer thickness we chose a design based on a conservative drive voltage (unbiased ac signal) of ±150 V. Therefore, an eight-wafer stack was required to meet the displacement requirement.

The wafers are stacked with every other layer flipped such that the positive (+) and the negative (−) surfaces are together. This allows the stack to be connected mechanically in series, but electrically in parallel. Hence the displacements generated by each layer are additive. The use of an even number of wafers is highly advantageous in control applications since this configuration provides substantial self-shielding from radiated EMI. The wafers were 2.54 cm square, and purchased with plated surfaces. They were assembled using 50 micron thick expanded nickel sheets between each layer, and glued using a low-viscosity epoxy under light pressure. The expanded metal provided an electrode connection to each wafer surface.

In addition to the primary actuator, two additional actuators were also fabricated from the "soft" and "single-crystal" classes of PZT to study the influence of actuator linearity on controllability and out-of-band enhancement. For the "soft" material we choose a PZT-5H type (Channel Industries, type C5700). Available wafers were 2.0 mm thick, and had a piezoelectric charge constant of 550 pm/V. This wafer thickness can be safely driven by an unbiased ac signal at our maximum drive level of 300 V. Hence a two-layer stack will provide a displacement of 0.33 microns, which satisfies our design requirement. The principle advantages of using this material are its ready availability and its ability to deliver the required displacement in a smaller form factor using only two PZT wafers. This also minimizes the number of glue bonds, and the issues associated with them. For the single-crystal material, we used lead magnesium niobate-lead titanate, PMN-32% PT wafers (TRS Ceramics, Inc.). This transducer material has only recently become available commercially. The displacements generated by these particular wafers were initially evaluated using a laboratory fiber optical displacement sensor (PhilTech Model RC12-A2OQR). These wafers were from early production runs, and had variations in their properties. A pair was selected for a two-wafer stack with an effective (combined) d_{33} of 2750 pm/V. When driven at 125 V this stack develops 0.344 microns displacement.

In implementation, a thin rubber layer (0.06 in.) layer was placed against one side of the actuator to fully realize the actuator's drive level.

B. Sensors

Some requirements for the force sensor are similar to those of the actuator, including static load survivability, low drift and small time delay. An important additional requirement is that its noise floor be sufficiently low so that accurate monitoring of the (suppressed) disturbance is possible. This requirement directly affects the level of achievable performance. The maximum dynamic force was previously indi-

cated as 1100 dynes. Because of the wide dynamic range anticipated for these signals, we desire 100 dB of S/N, or a noise floor of 0.011 dyne.

The force sensor was fabricated using piezoelectric materials. Since the sensor is typically connected to a high impedance preamplifier, the dielectric displacement (current drain) is negligible, and the inherent force sensitivity S_f (in volts per Newton) is

$$S_f = -g_{33}\frac{A}{z}, \tag{4}$$

where A and z are the area and thickness, and g_{33} is the piezoelectric voltage constant.

The preferred force sensor consists of two sensor layers wired as a shielded differential pair encased in a small η-metal Faraday cage. This shielding is helpful in reducing stray pickup of electromagnetic interference (EMI) from the actuator drive, which is an important consideration in active control systems. A variety of piezoelectric materials were considered. The expected noise floor for representative candidates of each material type was calculated, where the sensor is considered as just one component of a circuit that also includes the connecting leads and preamplifier. The model used is the equivalent network model of Tims[12] as applied in previous studies.[13] Comparing these modeling results, the sensor material chosen was type PR-305 piezorubber (NTK Piezoelectric Ceramics Division, NGK Spark Plug Co.) This material is readily available with silver ink electrodes. The sensor consists of two 1.25 mm thick layers of this material, forming a 6.45 cm^2 area sensor. The estimated sensitivity is 0.33 V/N, the capacitance is 270 pF, and the dissipation in this material is nominally 2%.

The key accelerometer requirements are low noise floor and freedom from extraneous interference. Mechanical considerations (i.e., static and shear load) are less restrictive than those of the other two components since the accelerometer is (typically) mounted in a force-free container or location. As with the force sensor, we assign the noise floor requirement as follows. The maximum acceleration at the 50 Hz test frequency was stated to be 0.003 g. Because of the wide dynamic range anticipated for these signals, we desire 100 dB of S/N, or a noise floor of approximately 0.3 μg.

This noise floor requirement is quite restrictive, and requires a very low noise accelerometer. For the tests described, we used Wilcoxon Model 759 accelerometers, which have a listed noise floor of 0.30 $\mu g/\sqrt{\text{Hz}}$. The size of these accelerometers (nominally 2 cm) is larger than can be easily embedded in the device under test. However for these tests, size was unimportant since the accelerometers could be located on structural elements, on the driven mass and on the supporting base. (For practical applications, smaller accelerometers with adequate noise floors can be fabricated.[14,15])

C. Rubber isolator

The rubber isolator sets the baseline (passive) mechanical impedance of the device, which largely defines the mechanical operating environment. The principle parameter governing the dynamics of this isolator is the effective spring constant,

$$k_e = E_{\text{eff}}\frac{A}{t}, \tag{5}$$

where A is the area of the loaded surface and t is the thickness. The effective modulus, E_{eff} is a function of both Young's and Bulk modulus and modified by the shape of the rubber component. For present purposes the effective modulus can be estimated using Snowdon's equation,[16]

$$E_{\text{eff}} = \frac{E(1+\beta S^2)}{1+\frac{E}{B}(1+\beta S^2)}, \tag{6}$$

where E and B are the Young's and Bulk modulus of the rubber. The parameter S is called the "shape factor" which is just the ratio of the area of one loaded surface to the total force-free area. For the case of a disk or block of material, this shape factor is

$$S_{\text{disk}} = \frac{r}{2h}, \quad S_{\text{block}} = \frac{L_a L_b}{2h(L_a+L_b)}, \tag{7}$$

where L_a and L_b are the dimensions of the loaded surface. The remaining parameter β, is a function of the hardness of the rubber. It can be approximated from the Shore-A (SA) hardness,

$$\beta = \begin{Bmatrix} \text{SA} > 70 & 1 \\ 20 < \text{SA} < 70 & 2.4 + 0.02\text{SA} \\ \text{SA} < 20 & 2 \end{Bmatrix}. \tag{8}$$

The rubber component in series with a mass M (kg) will then have a classical spring-mass resonance whose frequency is given by

$$\omega_r = \sqrt{\frac{k_e}{M}}, \quad \omega_r = 2\pi f_r, \tag{9}$$

where k_e is stiffness in N/m, f_r is the frequency in Hertz, and ω_r is the angular frequency in radians.

To minimize the height of this rubber layer, a very soft rubber was desired. The rubber used (Forsheda Seals Inc.) had a listed Shore-A durometer value of 10, but a measured value of 15. Its Shear modulus was determined using a torsional resonance measurement. Since the material is very soft, the value of Poison's ratio should be very close to 0.5, and the Young's modulus will be nominally three times the value of the Shear modulus. The characteristics of this rubber are given in Table I.

III. EXPERIMENTAL CONFIGURATION

The one-dimensional hybrid isolator experiments were conducted at the Laboratory for Vibro-acoustics, Naval Research Laboratory, Washington, DC. The laboratory configuration of the control experiment is shown in Fig. 2. The system has three main elements. The upper element includes a 5.6 kg mass and Wilcoxon F3/F9 electromechanical shaker located at the center of the mass. This shaker is typically

TABLE I. Rubber isolation pad dimensions and properties.

Durometer	15 (Shore A)
Area	200 mm^2
Height unloaded	4.57 mm
Height loaded (with 6.12 kg)	3.43 mm
Resonant frequency	19 Hz
Shear modulus (measured)	0.126 MPa
Young's modulus	
$3\times$ Shear	0.37 MPa
measured under load	0.45 MPa

driven with a sinusoidal signal at 50 Hz, and the drive signal level is maintained at 0.003 $g/\sqrt{Hz}$ to simulate the vibrating machinery. The second element is the hybrid device itself, which consists of a stacked PZT actuator, a passive compliant isolator, a force sensor, and two acceleration sensors. These two accelerometers are attached on the mass and base, along the center axis of the device. The third element is the flexible base system, which simulates to some degree the platform or deck that would be supporting the machinery. It has a mass of 12.6 kg and is supported on air isolation mounts (Enidine, Inc.) to prevent ground vibration from contaminating the experimental system. It also contains a performance accelerometer mounted away from the mount to measure the level of downstream system performance.

The experimental hybrid mount system forms a coupled system with two compliant elements (the passive device stage and the base plate air isolation mount.) The system can be modeled using a lumped parameter approach as shown in Fig. 3. The model is essentially a two degree-of-freedom system with two additional degrees-of-freedom added to model the piezoelectric actuator and force cell stage. The equations of motion for this system are

$$
\begin{aligned}
&-F_{\mathrm{PZT}}+F_D=M_1\ddot{x}_1+k^*_{\mathrm{PZT}}(x_1-x_2),\\
&F_{\mathrm{PZT}}=M_2\ddot{x}_2+k^*_{\mathrm{PZT}}(x_2-x_1)+k^*_{\mathrm{CI}}(x_2-x_3),\\
&0=k^*_{\mathrm{CI}}(x_3-x_2)+k^*_{\mathrm{FS}}(x_3-x_4),\\
&0=M_3\ddot{x}_4+k^*_{\mathrm{FS}}(x_4-x_3)+k^*_{\mathrm{GI}}x_4,
\end{aligned}
\tag{10}
$$

where F, M, k, and x refer to applied force, mass, stiffness, and displacement, respectively. The subscripts refer to the following: PZT, PZT actuator; D, disturbance; CI, compliant isolator element in hybrid mount; FS, force sensor; and GI, ground isolator (air mounts). The mass, M_1 is the load (5.6 kg) plus one-half of the PZT actuator mass. The mass, M_2 is the other half of the actuator mass, and M_3 is the mass of the base plate. Also shown in Fig. 3 is the equivalent circuit model of the system that is used to couple the electromechanical devices with the other mechanical elements of the system.

A comparison of the model results and laboratory data are shown in Fig. 4. Magnitude and phase displays of the base plate velocity are shown for the cases where the shaker disturbance and the PZT-5H control actuator are driven separately. Similar agreement was also found at the other accelerometer and force cell locations. The classical features of an isolator in a two-degree of freedom system can be observed. The two fundamental system modes occur at 12.0 and 25.6 Hz. The 12.0 Hz mode is due to the total mass above the air mounts and the stiffness of the air mounts. The second mode at 25.6 Hz is related to the stiffness of the passive element in the hybrid actuator and the mass on either side of it. The agreement between experiment and model indicates that the resonant motion and the mass/stiffness properties of the elements are well understood.

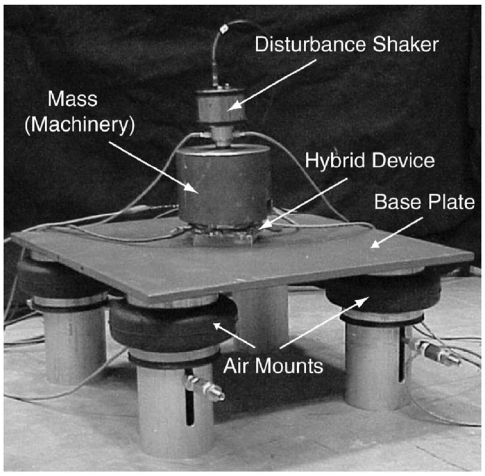


FIG. 2. Photograph of the one-dimensional test configuration.

The controller is based on a conventional adaptive (LMS) control algorithm, illustrated in Fig. 5. The electrical reference drives the disturbance shaker, and in turn, excites the structure. With the electrical reference and error sensor signal as input, the LMS algorithm adaptively adjusts the weights in the finite impulse response (FIR) filter until satisfactory cancellation occurs at the error sensor. The filter or-

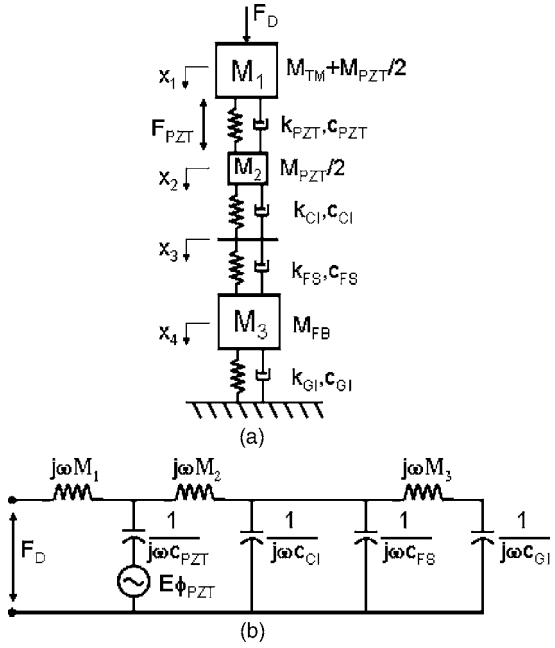


FIG. 3. Illustration of (a) lumped parameter model and (b) equivalent circuit.

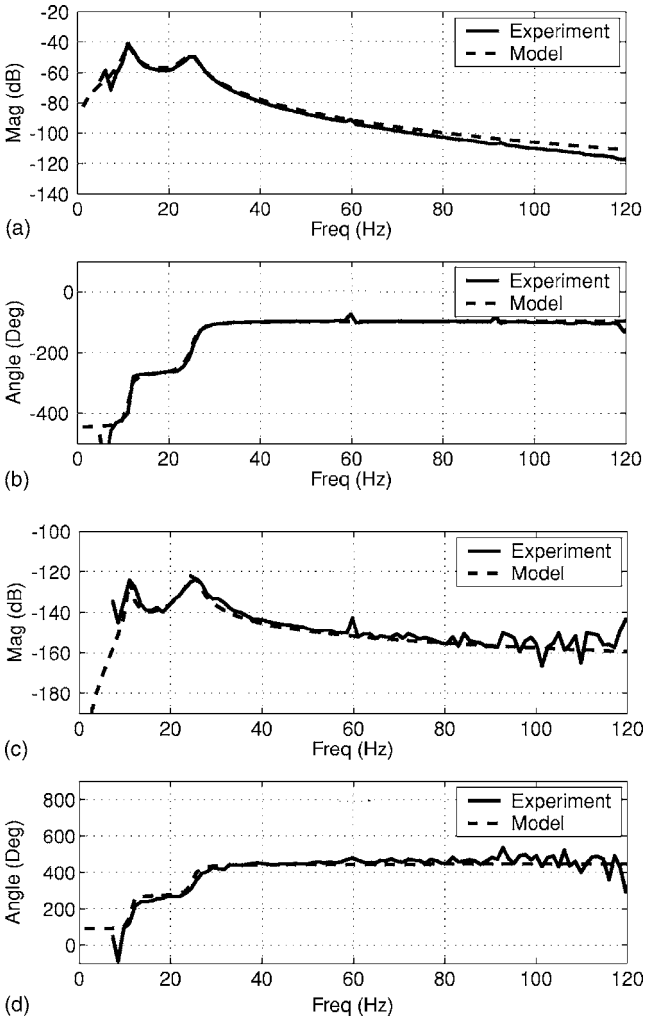


FIG. 4. Comparison of model and experimental base plate velocity results for the hybrid isolator. Displayed are (a) magnitude and (b) phase curves for the disturbance driven [(m/s)/N] case, and (c) magnitude, and (d) phase for the actuator driven [(m/s)/Volt] case.

der was determined by incrementally increasing the number of taps until little improvement was found in terms of the performance. Only a relatively simple, low order 8-tap FIR filter was required in these experiments.

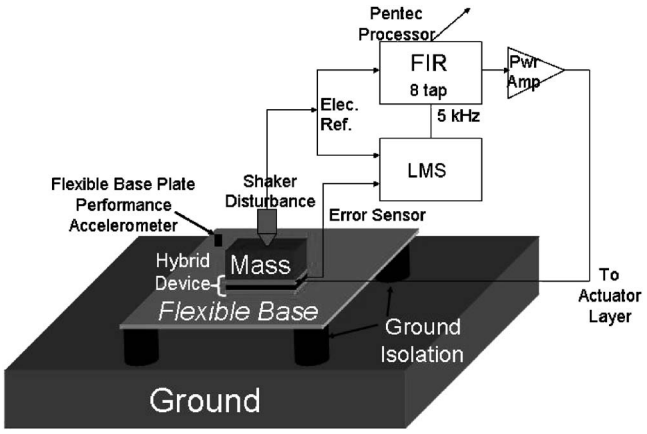


FIG. 5. Setup and implementation of the adaptive LMS/FIR based controller.

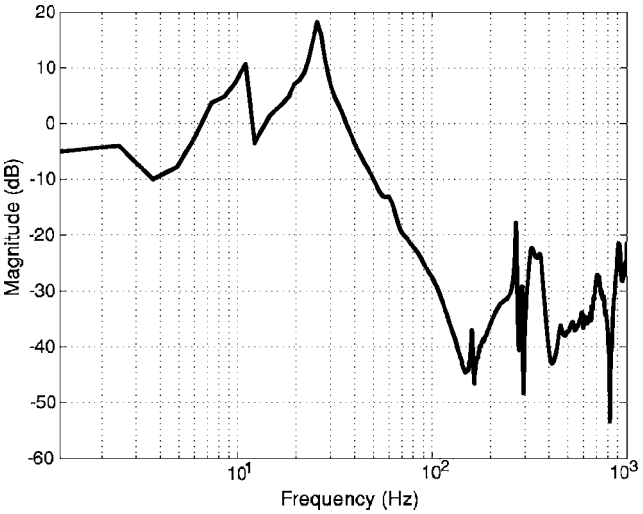


FIG. 6. Display illustrating passive performance. The transfer function is the base acceleration response with the passive rubber isolator over the response without the passive rubber isolator (mechanically shorted). Levels below 0 dB indicate the amount of dB reduction.

IV. HYBRID CONTROL STUDIES

A. Performance

The goal of this study is to reduce the amount of motion and associated mechanical power that is transmitted through the mount from the disturbance to the base plate. Two system performance metrics will be used to evaluate the device. The first metric is the input power to the base plate, or the power flow through the mount. The power flow may be calculated[17] as

$$P = |F_{\text{peak}}||v_{\text{peak}}|\cos(\phi), \tag{11}$$

where F is the measured force in the mount, v is the velocity (determined from an accelerometer measurement), and ϕ is the phase angle between F and v. The second metric is determined by a downstream performance accelerometer, mounted at a location away from the mount configuration. Other studies[2] have used energy based performance metrics with a high density of accelerometers covering the base plate structure. The approach in this paper is to use a small number of local measurements to determine the global system performance. The power flow metric involves local force and velocity measurements that are used to determine the power transmitted to the base plate system. This is done without knowledge of how the power is distributed in the plate afterward, but it does indeed yield a performance metric of the entire system without the requirement of a large matrix of sensors. The second metric uses a point acceleration measurement to obtain a global measure of the system. This is possible due to the design of the base plate system, where no plate modes that exist near the 50 Hz evaluation frequency and the base plate structure nominally moves as a rigid body.

The system performance is evaluated for three conditions (1) with no passive or active isolation (mount mechanically shorted), (2) with passive isolation only, and (3) in the hybrid (passive-active) configuration.

A transfer function for the passive mount (Fig. 6) shows

TABLE II. Hybrid actuator stacking configuration.

Stack 1A (APS)	Stack 1B (SAP)	Stack 2A (PAS)	Stack 2B (SPA)
Actuator layer	Force sensor	Passive isolator	Force sensor
Passive isolator	Actuator layer	Actuator layer	Passive isolator
Force sensor	Passive isolator	Force sensor	Actuator layer

that a reduction of typically 20–40 dB is found at frequencies above 75 Hz, or approximately 3 times the mount fundamental frequency. Also observed is the expected enhancement at very low frequencies associated with operation near the fundamental mount system resonances. At the 50 Hz test frequency of particular interest here, the level of reduction due to the passive element was 10.5 dB as measured by the force sensor, 10.2 dB as measured by the base accelerometer, and 9.6 dB measured downstream by the performance accelerometer.

The optimal active control law will be identified through experiment by determining the mount and controller configuration that yield the highest system performance. This goal will be achieved by exploring a matrix of (1) locations in the mount for error sensing, (2) locations in the mount for actuation, and (3) the error sensor quantity (force or velocity).

The six possible series configurations for the actuator (A), passive element (P), and force sensor (S) are APS, ASP, PAS, PSA, SAP, and SPA. Recall that the accelerometers are fixed at locations above and below the hybrid device. The first letter designation corresponds to the layer that is against the upper disturbance mass, the second letter is the element sandwiched between the first and third elements, and the third letter is the element against the lower base plate. The configuration pair of SAP-ASP involves redundancy because the force measured on either side of the actuator will be essentially identical due to the low mass and low loss of the actuator. Further, the hybrid elements above and below the passive element are non-wavebearing at low frequency, and they will move together at low frequency. Therefore, the acceleration can essentially be measured anywhere on the central axis above and below the soft rubber element. The same redundancy conclusion can also be drawn below the passive element for the PAS-PSA pair. Therefore, there are only four nonredundant configurations, shown in Table II. To simplify the designations, the stacks are put into two groups. The two stacks designated with 1 have the actuator located above the passive element, and the two stacks designated with 2 have the actuator below the passive element.

The performance of each of these stack arrangements was evaluated using the PZT-4 actuator. The adaptive LMS controller minimizes the signal at each error sensor (force cell, base accelerometer, and mass accelerometer) individually using +/−200 volts. The control measurements for the Stack 1A (APS) and Stack 2A (PAS) configurations each require three separate control studies (using each of the three error sensors). During each study, the error signal as well as all other available sensors was measured. The arrangement of Stack 1B (SAP) and 2B (SPA) differed from those of Stack 1A (APS) and 2A (PAS) only in the location of the force sensor. Hence, the only control study required with stack 1B and 2B is the use of the force sensor as the controller's error signal input.

An example of broadband response curves for the cases of no isolation, passive isolation, and hybrid isolation are shown in Fig. 7 due to the single frequency 50 Hz disturbance. The hybrid configuration is for the case where the base plate accelerometer is minimized by an actuator located above the passive mount element. The performance results for all the control experiments are summarized in Table III. The total system performance given in the table combines the results of both the active and passive mechanisms. The rows of the table correspond to separate control experiments, and the columns list the reductions for each of the sensors. For each experiment, the sensor used as the error signal is seen to have its amplitude reduced by the greatest amount when the hybrid controller is activated. (These values are identified using underlined italics in the table.) Of principle interest is the reduction found at the performance accelerometer and through the power flow calculation (first and last columns shown in bold). Recall that the performance accelerometer is viewed as a downstream metric that will accurately evaluate the level of isolation provided between the disturbance and the base support structure. The breakdown of the power flow calculation [Eq. (11)] reveals that the majority of the power flow performance is coming from magnitude reductions in both force and velocity. The phase angle contribution to the power flow performance level was found to average only 1–2 dB.

Examining both performance metrics for Stack 1 (actuator above), the best performance is found when the base accelerometer is used as the control error sensor. In the case of Stack 2 (actuator below), the downstream metric again showed that the best performance was with base acceleration minimization while the power flow metric indicated that force minimization (either above or below passive element) was better by ~6 dB. In the Stack 2 power flow calculation for the base accelerometer minimization experiment, there is significant reduction in the base accelerometer (61.6 dB), but the force (−1 dB) remains nominally unchanged. Whereas, in the force minimization experiment (Stack 2), the base accelerometer is reduced by 11–12 dB as well as having the force (62–63 dB) significantly reduced.

Both metrics, however, are consistent when examining the best overall system performance. This is achieved with the arrangement where the actuator is above and the accelerometer (base) error sensor is below the passive element. A power flow reduction of 54.7 dB and a 35.1 dB reduction at the downstream performance accelerometer are achieved in this optimal configuration. Note that the power flow reduction value is significantly larger due to the fact that power is determined through the multiplication of force and velocity, and not velocity alone as with the downstream metric. These performance levels are significantly higher than the nominal 10 dB of performance produced by the passive element alone.

Note that in almost all of the control studies, the levels below the mount are significantly reduced, but the above mount (mass accelerometer) levels have increased 8–10 dB.

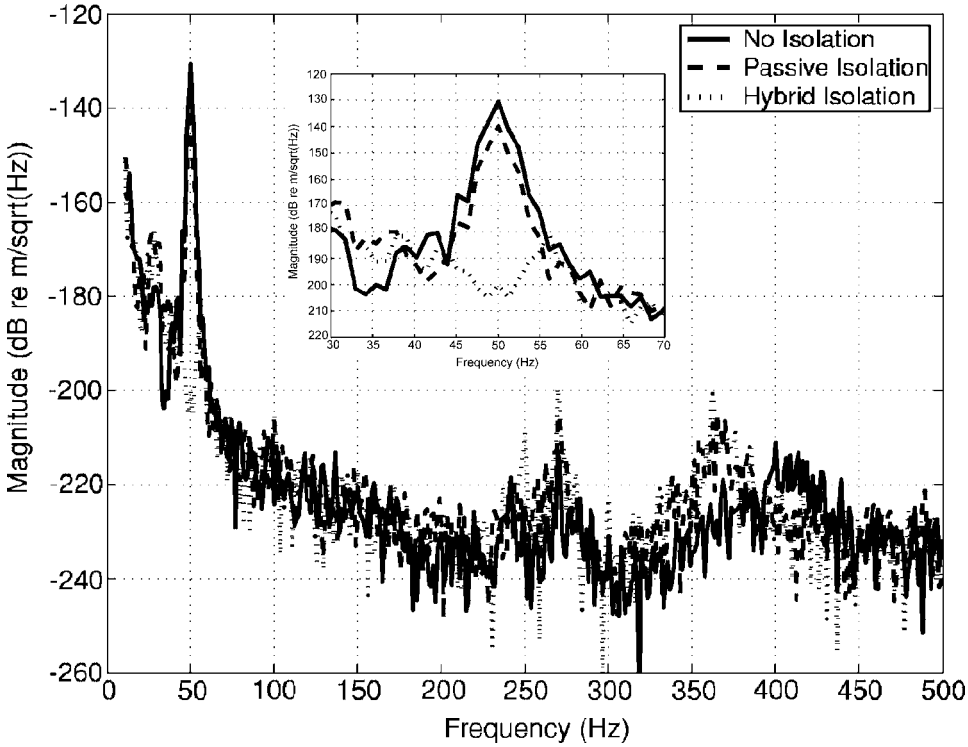


FIG. 7. Hybrid experimental result (meter/$\sqrt{\text{Hz}}$) at base accelerometer error sensor. The PZT4 actuator is located above the passive element, and base acceleration is minimization at 50 Hz. Passive performance of 10.2 dB and hybrid (active-passive) performance of 70.3 dB is found at base accelerometer relative to the no isolation case. Note the very low levels of harmonic distortion.

The only exception is for the trivial case where the mass accelerometer itself is the minimized error sensor. These increased levels at the mass are a reaction to the force generated by the actuator, and the active operation of the hybrid isolator as a more compliant support. These levels are deemed acceptable so long as they do not effect the functioning of the isolated machine.

While the above tests were all conducted at 50 Hz, some additional tests were also performed at frequencies above and below this key frequency to assess robustness. The total combined performance results (passive-active) at 25, 50, and 270 Hz were very similar, but the reductions at each frequency had different levels of contribution from the passive and active mechanisms. This is a result of the frequency dependence of the isolation provided by the passive component, and the associated reduction of signal-to-noise ratios (SNR) seen by the active system at higher frequencies. The passive component is very effective at 270 Hz, significantly

TABLE III. Combined (active-passive) performance results for the possible hybrid mount configurations using PZT4 actuator.

	Performance measured at sensor (dB)				
Error sensor in minimization experiment	Performance Accel.	Base Accel.	Mass Accel.	Force sensor	Power flow
ACTUATOR ABOVE THE PASSIVE ELEMENT					
Stack 1A	Force sensor located below active and passive element				
Base acelerometer	**35.1**	*70.3*	(−8.3)	32.4	**54.7**
Force sensor	**25.6**	28.7	(−8.4)	*67.2*	**47.7**
Mass accelerometer	(**−1.7**)	(−2.3)	*39.5*	(−3.1)	**−2.7**
Stack 1B	Force sensor located above active and passive element				
Force sensor	**11.2**	11.3	(−9.8)	*68.9*	**39.7**
ACTUATOR BELOW THE PASSIVE ELEMENT					
Stack 2A	Force sensor located below active and passive element				
Base accelerometer	**27.9**	*61.6*	(−7.9)	(−1.0)	**31.1**
Force sensor	**12.4**	12.4	(−9.6)	*62.4*	**37.1**
Mass accelerometer	(**−1.0**)	(−1.0)	*40.4*	(−15.8)	**−8.5**
Stack 2B	Force sensor located above active and passive element				
Force sensor	**22.9**	11.1	(−8.6)	*63.3*	**37.6**

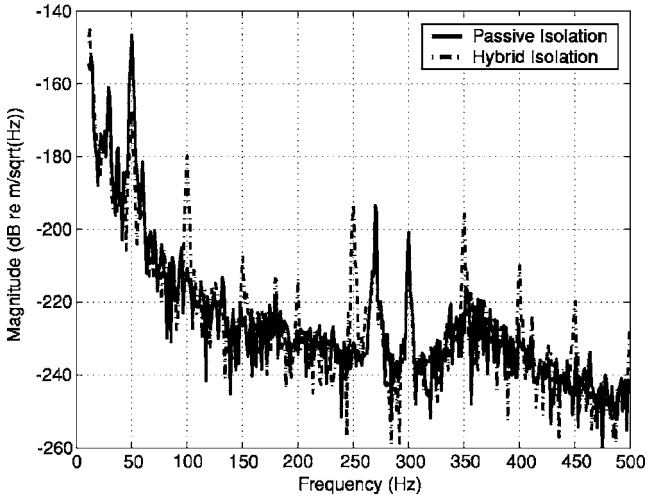


FIG. 8. Hybrid PZT-5H experiment result (meter/$\sqrt{\text{Hz}}$) at base accelerometer error sensor as in configuration of Fig. 7. Note the very high levels of harmonic distortion especially at the first, fourth, and sixth harmonics.

reducing the signal level and the associated SNR of the below mount error sensor signals. The controller can only suppress the tone to a limit associated with the sensor noise floor. At 25 Hz, on the other hand, the SNR levels are higher since the passive isolation is less effective, and the active component contributes to most of the performance attained.

B. Associated harmonic distortion

The influence of actuator nonlinearity on control and performance was also studied. Actuator nonlinearities generate harmonic distortion such as that shown in the data of Fig. 8 with a PZT5H control actuator. Recall that PZT5H is a soft material with higher levels of nonlinearity. In this example, the controller has minimized the targeted response at 50 Hz by ~25 dB more than the passive alone. However, higher order harmonics are excitation, creating very high levels of out-of-band enhancement.

One simple measure of nonlinearity is the total harmonic distortion (THD) of the system. THD is calculated here from the spectra by summing the significant harmonics and dividing by the amplitude of the fundamental peak. Eight harmonics were sufficient for this calculation since higher harmonics were always at least two orders of magnitude lower and typically near the noise floor of this measurement system.

During the performance control study, the values of THD were also determined at each sensor for each test configuration. The actuator delivers force and displacement to the system. Therefore, THD is calculated directly in terms of the force spectrum, and the acceleration outputs were integrated twice to obtain the displacement spectrum. The baseline level of system THD is 0.04% derived from the electrical source signal output of the L6 (Instruments Inc.) high voltage amplifier that was used to drive the PZT devices.

These THD values obtained from the control experiments using the PZT4 actuator are listed in Table IV. These values are relatively small, nominally less than 0.5%, which is considered suitable for control applications. On average, the actuator-above configurations have a THD value that is half (6.1 dB) of the actuator-below configurations. This difference is partially attributed to the frequency dependence of the passive isolator performance. Hence when the passive isolator is located between the actuator and sensor, the higher-frequency harmonics are reduced more than the fundamental.

It is noted that the force cell measurements (last column) usually showed significantly higher harmonic distortion than that associated with the accelerometers. This is probably in-

TABLE IV. Harmonic distortion results for possible hybrid mount configurations using PZT4 actuator.

	THD measured at sensor (%)			
Error sensor in minimization experiment	Performance Accel.	Base Accel.	Mass Accel.	Force Sensor
	ACTUATOR ABOVE THE PASSIVE ELEMENT			
Stack 1A	Force sensor located below active and passive element			
Base accelerometer	0.18	*0.04*	0.01	0.54
Force sensor	0.10	0.03	0.02	*0.64*
Mass accelerometer	0.25	0.10	*0.16*	1.25
Stack 1B	Force sensor located above active and passive element			
Force sensor	0.18	0.12	0.25	*0.55*
	ACTUATOR BELOW THE PASSIVE ELEMENT			
Stack 2A	Force sensor located below active and passive element			
Base accelerometer	0.36	*0.19*	0.08	2.10
Force sensor	0.04	0.03	0.01	*0.41*
Mass accelerometer	0.47	0.29	*0.40*	1.86
Stack 2B	Force sensor located above active and passive element			
Force sensor	0.40	0.42	0.08	*2.55*

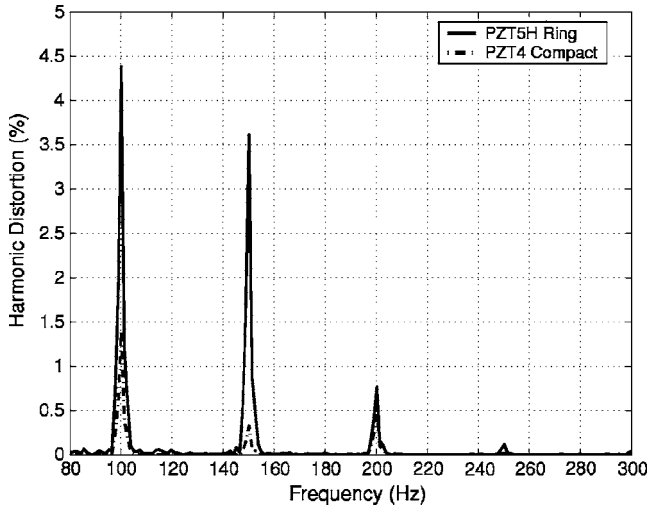


FIG. 9. Harmonic distortion associated with the PZT-4 and PZT-5H actuator material.

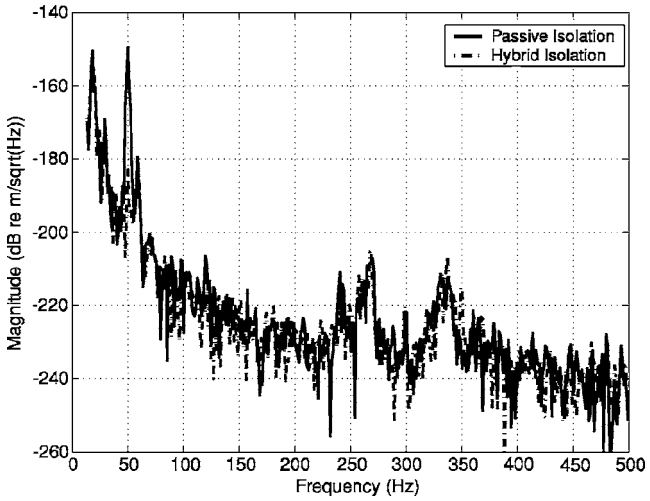


FIG. 10. A hybrid experiment result (meter/$\sqrt{\text{Hz}}$) at 50 Hz with PMN-32%PT single-crystal actuators. Base acceleration is again minimized as in the configuration of Fig. 7.

herent to the particular piezorubber force cell design used. Piezorubber has 2% inherent damping, and no precautions were taken to isolate strain along the strain axis.

Signal distortion components can be amplified by the controller and, if sufficiently large, can lead to stability limitations. Hence if distortion is a significant contributor to the performance limit, the optimum configuration would be expected to have the lowest THD as observed on the control error sensor. Examining the data in these tables, the levels indicated with underlined italics correspond to cases where the sensor is serving as the control error sensor. The lowest THD levels are seen to be the cases where the actuator is above the passive element and the lowest of these is when the base accelerometer is the error sensor. It is difficult to isolate factors in the performance limit, but this is the same trend found in the experimental performance data. Aside from harmonic distortion and out-of-band enhancement issues, system nonlinearity may be significant in determining higher levels of performance.

C. Other actuator materials

Harmonic distortion was also measured for each of three actuators fabricated, the "hard" PZT-4 stack, "soft" PZT-5H stack, and the PMN-PT single crystal stack. Figure 9 shows examples of the harmonic spectrum levels measured with the PZT-4 and PZT-5H actuators. The 50 Hz source signal is applied to the actuator under test, which is typically fixed at the base and has an accelerometer attached to its top surface. The output of the attached accelerometer was integrated twice to obtain a displacement spectrum, and the THD was calculated as described earlier. The PZT-5H material exhibits about 4 times higher THD than the PZT-4. From such spectra, the THD values for these actuators were determined, and the results are given in Table V. Table V also includes similar THD values reported in the literature by Royston[18] for 1–3 composites formed from these two materials and evaluated at various field values. Because of the different measurement and calculation techniques, Royston's nonlinearity values are not directly comparable to our THD values. The values are approximately one-third of that in this paper. However, Royston's data clearly shows similar trends with material type (PZT5H has ~4 times higher THD than PZT4) and also illustrates the dependence on field strength.

TABLE V. Actuator total harmonic distortion (from displacement).

	PZT-5H		PZT-4	
Field (V/mm)	This study	Ref. 1	This study	Ref. 1
20		⩽0.2%		⩽0.1%
100	8.9%	1.1%		0.2%
200		2.2%	2.2%	0.5%

The differences associated with harmonic distortion in the control experiments can be seen in Figs. 7 (PZT-4) and 8 (PZT-5H). The relatively high THD value for PZT-5H was expected since this material is known to have one of the highest nonlinearity coefficients in the PZT family. The lower value for PZT-4 was also expected since it is one of the more linear materials in this family; PZT-8 is only marginally better. It is noted that the THD values (from Fig. 9) were obtained using drive voltages comparable to those used in the control studies, since linearity (and THD) are known to be a function of field strength.

A control experiment with PMN-PT is shown in Fig. 10. The THD level calculated from this control experiment is 0.17%. When compared to the THD level in the PZT4 control experiments of Fig. 7 and Table IV, the PMN-PT material is found comparable to PZT4 in terms of the THD, yielding negligible out-of-band enhancement. It should be noted that these THD values are partially limited by the noise floor of the measurement and may actually be lower.

An interesting result here is the extremely small THD value for the PMN-PT single crystal material. While this material was expected to have good linearity at the low field condition used here, the exceptionally high degree of linearity measured was unexpected. An interesting side note is that, to a rough approximation, the values we measured for

THD are approximately equal to those of damping. The PMN-PT single crystal material also has a very high d_{33} constant, yielding a high drive authority. Essentially, this means that the actuator stack made of PMN-PT material can be very compact, only about 25% of the PZT4 stack height with the same drive capability and linearity.

V. CONCLUSIONS

Vibration isolation is significantly increased when incorporating active elements into the passive design. The synergy of the hybrid design works such that the active system dominates the level of performance at low frequencies near the fundamental resonance of the passive system, and the passive system is very effective at frequencies greater than ~3 times the fundamental frequency. In this study, the optimal configuration for the hybrid system was found to have the actuator above the passive element and to minimize the base acceleration. This arrangement yielded the highest level of performance for both the power flow and downstream acceleration performance metrics, and with the lowest average THD. Using the PZT4 actuator material and the LMS control algorithm (8-tap FIR), the combined hybrid performance measured values were 54.7 dB of power flow reduction and 35.1 dB of downstream motion reduction. The THD at all accelerometer locations was found to be no greater than 0.25% and sufficient for active implementation. This performance was significantly higher that the 10 dB of baseline performance provided by the passive element alone.

The hybrid approach was also found to be robust. It provided nominally the same level of total performance at all frequencies tested (25–270 Hz). It decreased when an actuator material with significantly higher nonlinearity was used (PZT-5H). However every system tested proved robust and delivered performance significantly better than the passive element alone. Further, the optimal configuration determined in this paper lends itself to a robust design since the actuator is itself effectively isolated from the base structure. This can be seen as a failsafe, protective design where the base system is protected from actuator instabilities, and in the event of complete active system shutdown, the hybrid system reverts to the passive-only configuration.

A promising finding is the performance obtained with the single-crystal PMN-PT actuator material. The reduced stack height associated with this material is critical in the design for applications where small shelf mounts are needed in electronics cabinets. The high linearity and high displacement output suggest that it is the material of choice in the development of the future three-axis hybrid isolator as well as in other control applications.

ACKNOWLEDGMENT

The authors would like to acknowledge Naval Sea Systems Command for their support.

[1] J. P. Den Hartog, *Mechanical Vibrations* (Dover, New York, 1985).
[2] E. E. Ungar, "Vibration isolation," in *Noise and Vibration Control*, edited by L. Beranek and I. Ver (Wiley, New York, 1992).
[3] M. D. Jenkins, P. A. Nelson, P. J. Pinnington, and S. J. Elliot, "Active isolation of periodic machinery vibrations," J. Sound Vib. **166**, 117–140 (1993).
[4] P. A. Nelson, M. D. Jenkins, and S. J. Elliott, "Active isolation of periodic vibrations," *Proceedings of Noise–Control 1987*, 1987.
[5] B. Widrow and S. D. Sterns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985).
[6] A. H. von Flotow, "An expository overview of active control of machinery mounts," *Proceedings of the 27th Conference on Decision and Control*, 1988.
[7] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, San Diego, 1997).
[8] J. H. Su, "Robust passive-active mounts for vibration isolation," Proc. SPIE, Smart Structures and Materials, Vol. **3672**, 144–155 (1999).
[9] Morgan Matroc Inc., *Piezoelectric Technology Data for Designer* (Morgan Matroc Inc., Bedford, Ohio).
[10] See, for example, *Piezoelectric Ceramics* (Channel Industries, Inc., Santa Barbara, CA).
[11] K. G. Eyster, "Comparison of thermal and electric field test data for piezoelectric ceramic materials," NUWC-NPT Tech. Memo., Dec., 2000.
[12] A. C. Tims, "Hydrophone preamplifier optimization prediction of hydrophone self-noise by a noise model," NRL Report 8180, 1978.
[13] R. D. Corsaro, B. H. Houston, and J. A. Bucaro, "Sensor-actuator tile for underwater surface impedance control studies," J. Acoust. Soc. Am. **102**, 1573–1581 (1997).
[14] R. D. Corsaro and J. Klunder, "Monolithic piezoelectric accelerometers," US Patent No. 5869762, 1999.
[15] R. D. Corsaro *et al.*, "Co-formed accelerometer array for integrated sensor/actuator applications," *Acoustic Particle Velocity Sensors: Design, Performance and Applications*, AIP Conference Proceedings 368 (AIP Press, Woodbury, NY, 1995).
[16] J. Snowdon, "Vibration Isolation: Use and characterization," J. Acoust. Soc. Am. **66**, 1245–1274 (1979).
[17] I. Ver, "Interactions with solid structures," in *Noise and Vibration Control Engineering*, edited by L. Beranek and I. Ver (Wiley, New York, 1992).
[18] T. J. Royston and B. H. Houston, "Modeling and measurement of nonlinear dynamic behavior in piezoelectric ceramics with application to 1-3 composites," J. Acoust. Soc. Am. **104**, 2814–2827 (1998).

A numerical investigation of the influence of windscreens on measurement of sound intensity[a)]

Peter Juhl[b)]
Physics Department, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Finn Jacobsen[c)]
Acoustic Technology, Ørsted•DTU, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

(Received 9 February 2005; revised 12 September 2005; accepted 23 November 2005)

Sound intensity probes are often used with windscreens to minimize the effect of noise caused by airflow. A theoretical and experimental study of the effect of windscreens on *p-p* intensity probes published ten years ago concluded that windscreens give rise to underestimation of the sound intensity at low frequencies in strongly reactive sound fields. The theoretical part of this study was based on the assumption of a windscreen of infinite extent. In this paper windscreens of realistic size and shape are dealt with by means of a coupled boundary element model for the windscreen and the surrounding air. The error of the estimated intensity caused by the windscreen is calculated under a number of sound field conditions of varying reactivity. It is shown that the resulting error can be much larger than the intensity itself in a very reactive sound field. It is also shown that the shape and size of the windscreen has a significant influence on the error. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151811]

PACS number(s): 43.58.Fm, 43.50.Yw [DKW] Pages: 937–942

I. INTRODUCTION

Windscreens of porous foam are often placed around sound intensity probes to reduce their exposure to airflow and mechanical or thermal damage.[1–3] However, windscreens have an undesired effect; a theoretical and experimental study of the effect of windscreens on sound intensity measurements with the two-microphone technique has shown that the use of a windscreen results in underestimation of the sound intensity at low frequencies in strongly reactive sound fields.[4] The reason is that the losses of the foam gives rise to a phase error between the pressure and the particle velocity.[4] The theoretical part of this study was based on the somewhat oversimplified assumption of a windscreen of infinite extent, and some of the experimental observations indicated a significant influence of the shape of the windscreen that evidently cannot be predicted by a theory that assumes that it is infinite. Thus the purpose of this paper is to examine the influence of windscreens of realistic size and shape on intensity measurements numerically.

II. THEORY

A. Two coupled domains

Consider a windscreen placed in an arbitrary sound field. This may be regarded as a problem that involves two coupled domains, the exterior sound field and the field inside the windscreen. In the domain exterior to the windscreen the air is assumed to be an ideal fluid. If the sound field is harmonic the velocity potential satisfies the Helmholtz equation

$$\nabla^2\Psi_e + k_e^2\Psi_e = 0, \tag{1}$$

where k_e is the wave number, which is real valued in this case. The sound pressure and the particle velocity are

$$p_e = j\omega\rho_e\Psi_e, \tag{2}$$

and

$$\mathbf{u}_e = -\nabla\Psi_e, \tag{3}$$

respectively, where ρ_e is the density of air. Note that the $e^{j\omega t}$ convention is used in this paper.

The windscreen is assumed to be a porous medium as described by the simple phenomenological model of Zwikker and Kosten[5] and Morse and Ingard.[6] Such a medium is characterized by its porosity Ω_i, compressibility κ_i, density ρ_i, and flow resistivity Φ_i. In the following it is assumed that $\Omega_i = 1$, that the density equal the corresponding value of ideal air, and that the compressibility equals the isothermal compressibility of air, so that the speed of sound

$$c_i = 1/(\Omega_i\rho_i\kappa_i)^{1/2}, \tag{4}$$

is about 15% less than the speed of sound in the exterior domain c_e. It should be mentioned that many other, more accurate models of the acoustic properties of porous materials are available; see, e.g., Refs. 7 and 8. However, windscreens are made of highly porous foam with a low flow resistivity; their undesired effect on sound intensity measurement is negligible except at low frequencies;[4] and it is known that the Zwikker and Kosten model is a "phenomenologically correct low-frequency approximation."[8]

[a)]Portions of this work were presented in "A numerical investigation of the influence of windscreens on sound intensity measurements," Proceedings of 18th International Congress on Acoustics, Kyoto, Japan, April 2004.
[b)]Electronic mail: pmjuhl@fysik.sdu.dk
[c)]Electronic mail: fja@oersted.dtu.dk

In what follows indices of quantities that have the same values in the two domains have been omitted. The single remaining parameter in the model of the windscreen is the flow resistivity Φ_i, which enters into the equation of motion as follows:[6]

$$j\omega\rho\mathbf{u}_i + \Phi_i\mathbf{u}_i + \nabla p_i = 0. \tag{5}$$

The wave equation takes the form[6]

$$\nabla^2\Psi_i + (\omega/c_i)^2\Psi_i - \frac{j\omega\Phi_i}{\rho c_i^2}\Psi_i = 0, \tag{6}$$

and this can be transformed into the Helmholtz equation

$$\nabla^2\Psi_i + k_i^2\Psi_i = 0, \tag{7}$$

by defining a complex wave-number k_i as

$$k_i = \frac{\omega}{c_i}\sqrt{1 - \frac{j\Phi_i}{\omega\rho}}. \tag{8}$$

B. Coupled boundary element model

Since the exterior as well as the interior problem may be transformed into the regular Helmholtz equation, a standard Boundary Element Method (BEM) is employed for the numerical solution of the coupled problem.[9] The wave number of the interior problem is complex, but since its imaginary part is relatively small no special care is needed in this respect. (A large imaginary part would require special care in the numerical integration, due to the rapid decrease of the exponential.) The direct collocation method is used with three-dimensional quadrilateral, isoparametric linear elements. Two different windscreens have been modeled, the Brüel and Kjær (B&K) windscreen UA 0781, which is shaped like a prolate ellipsoid, and B&K's spherical windscreen UA 0782. The ellipsoidal windscreen is 18.5 cm long and has a width of 9.3 cm, and the spherical windscreen has a diameter of about 9.3 cm. Both models consist of 384 elements and 386 nodes, since the mesh for the ellipsoidal windscreen is produced by a simple transformation of the spherical mesh. The maximum distance between two nodes is 18 mm for the ellipsoidal windscreen, which corresponds to about 19 nodes per wavelength in air at the highest frequency considered in this study, 1000 Hz. In view of the normal rule of thumb of six nodes per wavelength and the fact that the effects studied in this work are most significant at frequencies below 100 Hz, it is concluded that the numerical mesh is more than adequate. It is also worth mentioning that the problem of characteristic frequencies[10] does not occur in coupled problems (and for an uncoupled problem the problem would occur at much higher frequencies, due to the size of the windscreens).

Once the exterior problem is discretized it becomes a matrix equation,

$$\mathbf{C}_e\mathbf{\Psi}_e = \mathbf{A}_e\mathbf{\Psi}_e + \mathbf{B}_e\mathbf{u}_e + 4\pi\mathbf{\Psi}_e^{\mathrm{I}}, \tag{9}$$

which can be converted to a matrix equation for the sound pressure $\mathbf{p}_e$ by multiplying with $j\omega\rho$,

$$\mathbf{C}_e\mathbf{p}_e = \mathbf{A}_e\mathbf{p}_e + j\omega\rho\mathbf{B}_e\mathbf{u}_e + 4\pi\mathbf{p}_e^{\mathrm{I}}. \tag{10}$$

In Eq. (10) $\mathbf{C}_e$ is a diagonal matrix containing the solid angle occupied by the calculation domain at the nodes (i.e., 2π for a smooth boundary), the matrix $\mathbf{A}_e$ contains integrals over the windscreen surface S involving the normal derivative of the Green's function, and the matrix $\mathbf{B}_e$ contains surface integrals involving the Green's function. The incoming sound field at position P is $\mathbf{p}_e^{\mathrm{I}}$ in the absence of the windscreen.

The interior problem can be converted to a matrix equation similar to Eq. (9), which is translated into the sound pressure $\mathbf{p}_i$ by multiplying with $j\omega\rho+\Phi_i$,

$$\mathbf{C}_i\mathbf{p}_i = \mathbf{A}_i\mathbf{p}_i + (j\omega\rho + \Phi_i)\mathbf{B}_i\mathbf{u}_i. \tag{11}$$

For both equations, Eqs. (10) and (11), the normal vector is defined so it points into the computational domain. Hence, the two domains share the surface S in the Helmholtz integral equation, but with opposite normal vectors. The diagonal matrix $\mathbf{C}_i$ contains the solid angles at all nodes measured from the interior domain; hence, this matrix complements the $\mathbf{C}_e$ matrix so that the two matrices add up to a diagonal matrix with 4π in all diagonal elements.

Continuity of the sound pressure and the particle velocity on the boundary of the windscreen leads to two equations

$$\mathbf{p}_e = \mathbf{p}_i = \mathbf{p}, \tag{12}$$

and

$$\mathbf{u}_e = -\mathbf{u}_i, \tag{13}$$

where the sign in the latter case is due to the change in normal direction on S (when considering the exterior and interior problem respectively). Rearranging Eq. (11) gives

$$\mathbf{u}_i = \frac{-1}{(j\omega\rho + \Phi_i)}\mathbf{B}_i^{-1}(\mathbf{A}_i - \mathbf{C}_i)\mathbf{p}, \tag{14}$$

and entering this result into Eq. (10) leads to a matrix equation for the pressure on the surface of the windscreen

$$\left(\mathbf{A}_e - \mathbf{C}_e + \frac{j\omega\rho}{j\omega\rho + \Phi_i}\mathbf{B}_e\mathbf{B}_i^{-1}(\mathbf{A}_i - \mathbf{C}_i)\right)\mathbf{p} = -4\pi\mathbf{p}_e^{\mathrm{I}}. \tag{15}$$

To summarize, Eq. (15) is the coupled equation for the sound pressure on the surface of the windscreen involving both the interior domain (the windscreen) and the exterior domain (the surrounding air). The matrices involving the exterior problem (denoted by index e) are standard BEM matrices with a real wave number k_e, whereas the BEM equations for the interior problem (denoted by index i) take the losses due to the flow resistivity into account by means of a complex wave number k_i.

Two limiting cases of Eq. (15) are of interest. If the flow resistivity Φ_i tends toward infinity the windscreen becomes in effect impenetrable [see Eq. (5)]. In this case the third term in the bracket on the left-hand side of Eq. (15) tends toward zero, and the problem reduces to the familiar boundary element formulation of the exterior problem of solving scattering from a rigid object $(\mathbf{A_e}-\mathbf{C_e})\mathbf{p_e}=-4\pi\mathbf{p_e^I}$. Conversely, if the flow resistivity vanishes (i.e., if the windscreen is removed), $k_i=k_e$, and therefore $\mathbf{B}_i=\mathbf{B}_e$ and $\mathbf{A}_i=-\mathbf{A}_e$ (where the change of sign in the latter expression is due to the

change of normal vector from the interior to the exterior problem), since the interior and exterior problems involve the same surface. Hence, Eq. (15) reduces to $(-\mathbf{C_e}-\mathbf{C_i})\mathbf{p_e}=-4\pi\mathbf{p_e^I}$ or $\mathbf{p_e}=\mathbf{p_e^I}$, since $\mathbf{C_e}+\mathbf{C_i}=4\pi\mathbf{I}$ (where $\mathbf{I}$ is the unitary matrix). Thus if the flow resistivity is negligible the windscreen does not change the sound field.

Once Eq. (15) is solved for $\mathbf{p},\mathbf{u}_i$ can be found using Eq. (14), and the pressure inside the windscreen can be found using Eq. (11) with P placed at any location inside the windscreen. If P is strictly inside the windscreen then $C_i(P)=4\pi$.

An intensity probe inside the windscreen could be modeled as described in Ref. 11. However, since windscreens have no influence on sound intensity measurements except at fairly low frequencies[4] where the probe does not change the sound field at all,[11] the intensity is simply calculated from the pressure at two points that represent the two microphones, that is, from the expression

$$\hat{I}_r=\frac{\mathrm{Im}\{p_1p_2^*\}}{2\omega\rho\Delta r}, \tag{16}$$

where Δr is the length of the distance between the microphones.

C. Infinite windscreen

If the windscreen is infinitely large its effect on sound intensity measurements can be derived analytically.[4] The particle velocity follows from Eq. (5)

$$u_{ir}=-\frac{1}{j\omega\rho}\frac{\partial p_i}{\partial r}\frac{1}{1-j\Phi_i/\omega\rho}. \tag{17}$$

However, the intensity estimate is based on a finite difference approximation to the equation of motion for a lossless medium

$$\hat{u}_r=-\frac{1}{j\omega\rho}\frac{\partial p}{\partial \mathrm{r}}, \tag{18}$$

that is

$$\hat{u}_r=u_{ir}(1-j\Phi_i/\omega\rho), \tag{19}$$

from which it follows that the estimated intensity is

$$\hat{I}_{ir}=1/2\ \mathrm{Re}\{p_i\hat{u}_r^*\}=\mathrm{Re}\left\{(I_{ir}+jJ_{ir})\left(1+\frac{j\Phi_i}{\omega\rho}\right)\right\}$$

$$=I_{ir}-\frac{\Phi_i}{\omega\rho}J_{ir}=I_{ir}\left(1-\frac{\Phi_i}{\omega\rho}\frac{J_{ir}}{I_{ir}}\right), \tag{20}$$

where J_{ir} is the r component of the reactive sound intensity[1,12]

$$\mathbf{J}=1/2\ \mathrm{Im}\{p\mathbf{u}^*\}. \tag{21}$$

The reactive intensity describes the nonpropagating part of the acoustic energy that is merely flowing back and forth.[12] Many sources have strongly reactive near fields at low frequencies where they essentially generate evanescent waves.[1,13] In general the reactive intensity points out of a source, corresponding to the radiation impedance being masslike. As can be seen from Eq. (20), the infinite windscreen theory predicts a negative bias error that is proportional to the reactive intensity and to the flow resistivity and inversely proportional to the frequency.

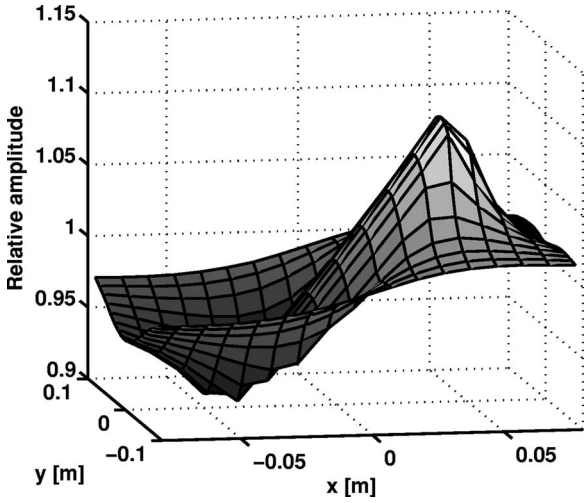


FIG. 1. Effect of spherical windscreen on normalized sound pressure amplitude in the x-y plane. The center of the windscreen is at $(0,0,0)$, and the sound field is generated by a monopole at $(0.10\ \mathrm{m},0,0)$.

III. NUMERICAL AND EXPERIMENTAL RESULTS

The coupled BEM model described above has been implemented in MATLAB, and some calculations have been made. Both the spherical and the ellipsoidal windscreen are made of the same material: open-pored polyurethane foam. The flow resistivity of a sample of the material has been measured using the two-microphone method described in Ref. 14 and found to be about 600 $\mathrm{kgm^{-3}\,s^{-1}}$.

Figure 1 shows the calculated amplitude of the sound pressure inside and just outside the spherical windscreen when a monopole emitting sound at 63 Hz is placed at $(x,y,z)=(0.10\ \mathrm{m},0,0)$, that is, 10 cm from the center of the screen, which has a flow resistivity of 600 $\mathrm{kgm^{-3}\,s^{-1}}$. The amplitude has been normalized with the distance to the source so as to have a value of unity everywhere in the absence of the windscreen. It is apparent that the pressure amplitude is increased by more than 7% by the windscreen near the monopole and reduced by a similar fractional amount near the opposite side. However, these modifications should be compared with the much larger pressure variations due to the varying distance to the source.

Figures 2(a) and 2(b) show the calculated phase of the pressure relative to the phase of the volume velocity of the monopole, with and without the spherical windscreen. Close to the monopole the phase angle approximates 90°. As can be seen the windscreen has a significant influence on the phase; in the middle of the screen the sign of the phase gradient is changed. Since the active intensity is proportional to the phase gradient of the pressure[12] this indicates that the measured sound intensity points in the wrong direction.

Some experiments have also been carried out. A B&K sound intensity probe of type 3599 with microphones of type B&K 4181 and the two windscreens UA 0781 and UA 0782 in combination with a B&K "Pulse" analyzer of type 3560 in

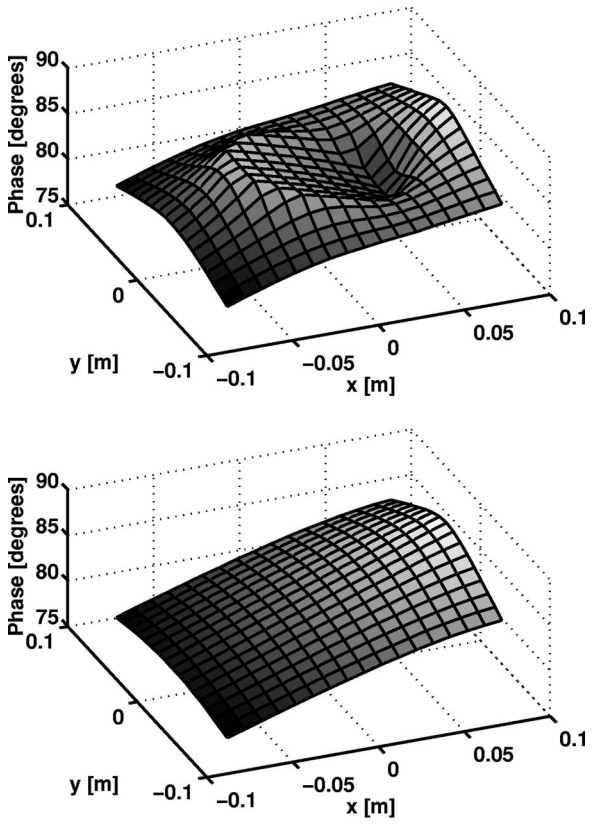


FIG. 2. Phase angle of sound pressure in the x-y plane relative to the volume velocity of the monopole that generates the sound field (a) with and (b) without a spherical windscreen centered at $(0,0,0)$. The monopole is at $(0.10\ \mathrm{m},0,0)$.

one-third octave mode were used in the measurements, which took place in DTU's large anechoic room. This room is very good at frequencies down to 50 Hz.[15] The first measurement was made near a "monopole," a small hole in a large baffle driven by an enclosed loudspeaker on the other side of the baffle; see Fig. 3. Figure 4(a) shows the sound intensity, the reactive intensity, and the particle velocity measured 11 cm from this source using a 12-mm spacer between the two microphones, without a windscreen on the probe and with the two windscreens. The active and reactive intensities have been normalized by $p_{\mathrm{rms}}^2/\rho c$ (where p_{rms} is the rms pressure without a windscreen on the probe), and the particle

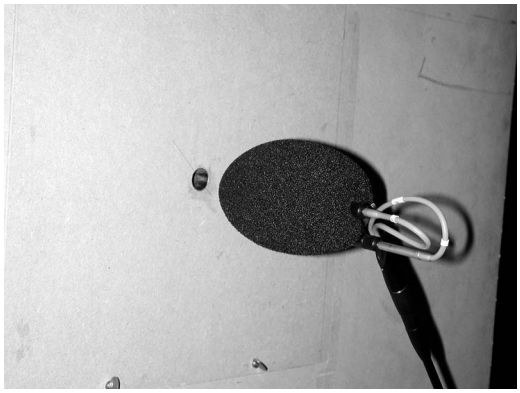


FIG. 3. Measurement near a baffled "monopole" with the ellipsoidal windscreen on the intensity probe.

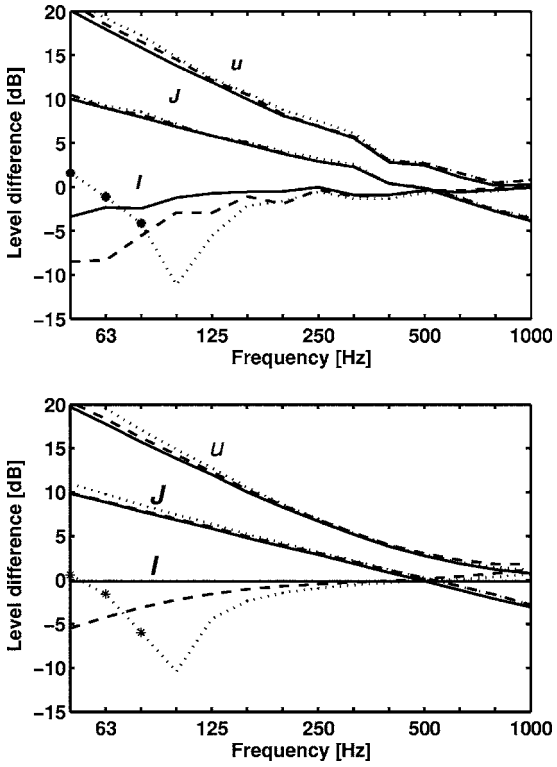


FIG. 4. (a) Measurements and (b) BEM calculations without windscreen (solid line), with an ellipsoidal windscreen (dashed line), and with a spherical windscreen on the intensity probe (dotted line) at a position 11 cm from a monopole. The stars indicate negative intensity.

velocity has been normalized by $p_{\mathrm{rms}}/\rho c$. Neither windscreen has any significant influence on the measured sound pressure (not shown) or on the particle velocity and reactive intensity. On the other hand the ellipsoidal screen has a considerable influence on the measured active intensity at low frequencies, and the spherical screen has a dramatic influence: It changes the sign of the intensity below 100 Hz. Figure 4(b) shows the results of BEM calculations for the same configurations with $\Phi_i = 600\ \mathrm{kgm^{-3}\,s^{-1}}$ and $\Delta r = 12$ mm. Considering the difficulty of constructing a "real monopole" there is very good agreement between measurements and calculations. Similar measured and calculated results (not shown) have been obtained with $\Delta r = 50$ mm at a distance of 13 cm from the source, with and without the ellipsoidal windscreen (the spherical windscreen cannot accommodate a 50-mm spacer). No influence of the spacer length was observed.

Another measurement has been carried close to a "dipole" constructed by mounting two loudspeaker units against each other. The moving part of this source, which is shown in Fig. 5, resembles a spherical source of first order (an oscillating sphere), and the reactivity (the ratio of the reactive to the active intensity) at a position on the axis of an oscillating sphere is[6]

$$\frac{J_r(r)}{I_r(r)} = \frac{\mathrm{Im}\{p(r)u_r^*(r)\}}{\mathrm{Re}\{p(r)u_r^*(r)\}} = \frac{1}{kr} + \frac{2}{(kr)^3}. \tag{22}$$

Figure 6(a) shows the measured reactivity at a distance of 21 cm from the loudspeaker dipole compared with the reactivity the same distances from a spherical source of first order. There is fairly good agreement (except perhaps at

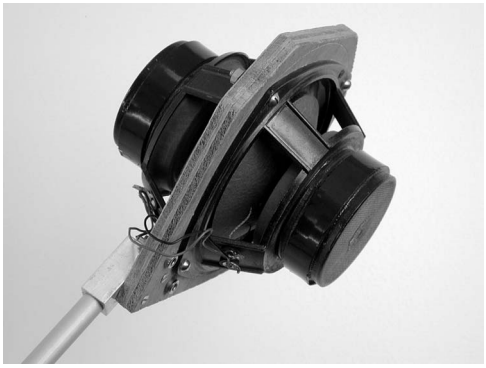


FIG. 5. The loudspeaker dipole.

80 Hz), confirming that the loudspeaker dipole can be modeled reasonably well as such a source (although less well than the experimental monopole described above can be modeled as a real monopole). Figure 6(b) shows the intensity measured with the spherical windscreen on the probe, normalized with the intensity measured without windscreen; the results of BEM calculations with the center of a windscreened intensity probe 21 cm from the center of a spherical source of first order using a value of 600 $\mathrm{kgm^{-3}\,s^{-1}}$ for the flow resistivity; and predictions based on the infinite windscreen theory [Eq. (20)] using the measured reactivity and a value of the flow resistivity found by adjusting it so as to get the best fit. Figure 6(c) shows similar measured and calculated results for the ellipsoidal windscreen. Note (again) that the spherical windscreen has a more serious influence on measurement of sound intensity than the larger ellipsoidal screen; the two windscreens change the sign of the measured intensity below 160 and 125 Hz, respectively. At 50 Hz the measured intensity with the spherical windscreen on the probe is no less than 50 times (17 dB) larger than the "true" intensity and pointing in the opposite direction. The influence of the two windscreens is predicted reasonably well by the BEM model considering the fact that the loudspeaker dipole is not a spherical source of first order. The agreement between the measured results and the predictions based on the infinite windscreen theory, Eq. (20), is also fairly good, in particular with the ellipsoidal windscreen (except at 100 Hz). However, rather different values of the parameter Φ_i were used in the two cases: 30 $\mathrm{kgm^{-3}\,s^{-1}}$ for the ellipsoidal windscreen and 80 $\mathrm{kgm^{-3}\,s^{-1}}$ for the spherical windscreen—although they are made of the same porous material and although the actual flow resistivity of the foam is much higher, about 600 $\mathrm{kgm^{-3}\,s^{-1}}$. In short, the parameter Φ_i in Eq. (20) is *not* the actual flow resistivity of the material but a quantity that also depends on the shape and size of the windscreen.

Finally Fig. 7 shows the results of a measurement with the ellipsoidal windscreen on an intensity probe in a standing wave tube with a standing wave ratio of 24 dB. The tube has been constructed for testing sound intensity measurement systems as specified in the IEC and ANSI standards.[16,17] The tube is 6.2 m long and has an inner diameter of 29 cm; see Ref. 18 for further details. The results shown in Fig. 7(a) have been obtained at 250 Hz with an intensity probe corrected for *p-p* phase mismatch and equipped with a 50 mm spacer.[18] As expected the largest deviations from the "true," constant intensity level occur midway between sound pressure maxima and minima where the reactive intensity assumes its largest values.[19] A prediction calculated with the coupled BEM model is shown in Fig. 7(b), where the amplitude of the incoming wave has been adjusted to match the level in the experiment. In the experiment the sound field at the probe was changed by changing its position in the duct, whereas the same effect has been obtained in the BEM calculations by changing the phase of the reflecting material at one end of the duct. There is very good agreement between the measurements and the BEM calculations.

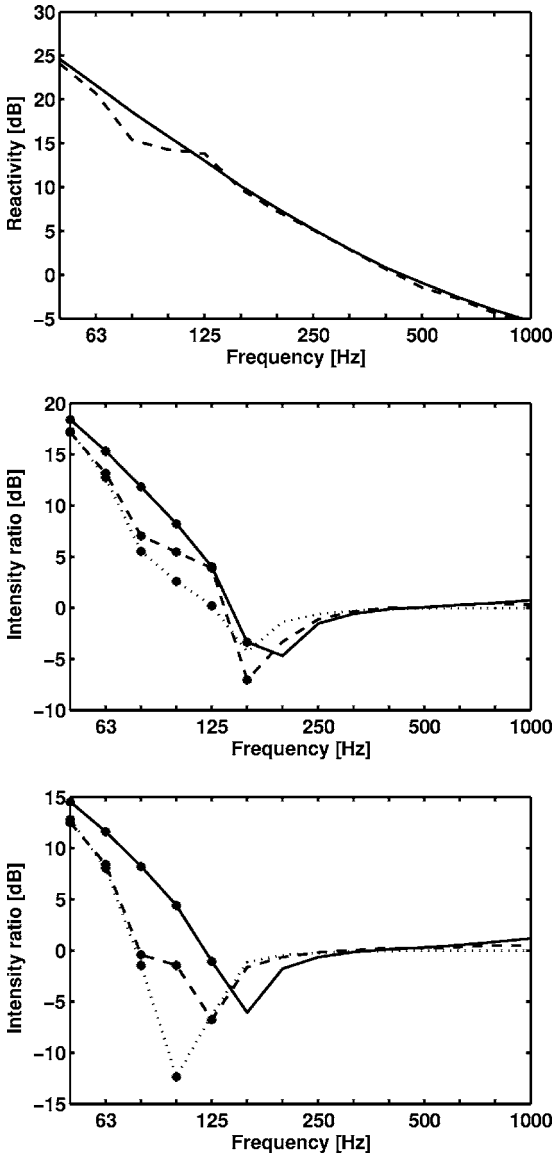


FIG. 6. (a) Reactivity, measured at a distance of 21 cm (dashed line) from the center of the loudspeaker dipole, and calculated 21 cm (solid line) from the center of a spherical source of first order; ratio of estimated to "true" intensity with (b) a spherical windscreen and (c) an ellipsoidal windscreen on the probe; measurements (dashed line), BEM calculations (solid line), and infinite windscreen theory (dotted line). The stars indicate negative intensity.

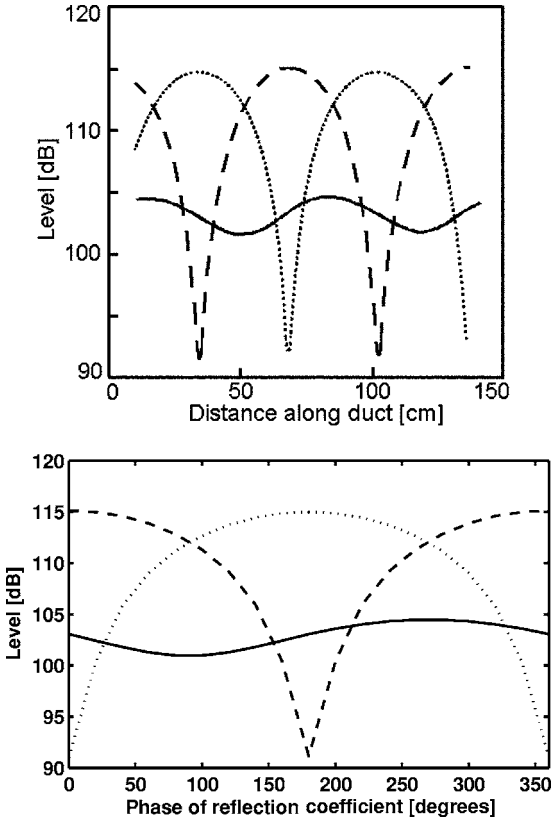


FIG. 7. Measured (a) and predicted (b) levels of sound pressure (dashed line), and particle velocity (dotted line) and intensity (solid line) in a standing wave with a standing wave ratio of 24 dB. Ellipsoidal windscreen with a flow resistivity of 600 $kgm^{-3}\ s^{-1}$.

IV. CONCLUSIONS

The effect of windscreens of porous foam on sound intensity measurements with the conventional two-microphone method has been examined numerically using a coupled BEM model. The results, which have been validated experimentally, confirm that windscreens give rise to significant bias errors, but only at low frequencies and in strongly reactive sound fields. Under such circumstances the error can be much larger than the intensity itself. By comparison windscreens have almost no effect on measurements of the sound pressure, the particle velocity, and the reactive intensity. The length of the spacer between the microphones does not matter. The size and shape of the windscreen are shown to be important: A large ellipsoidal windscreen gives much smaller errors than a smaller spherical one.

ACKNOWLEDGMENTS

The authors would like to thank Erling Frederiksen for data obtained with a windscreened sound intensity probe in Brüel and Kjær's standing wave tube for IEC 61043 tests.

[1]F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995).
[2]S. Gade, K. B. Ginn, and R. Hemelrijk, "Sound intensity measurements at 100 km/hour," *Proc. Inter-Noise 85*, Munich, Germany (Noise Control Foundation, Poughkeepsie, NY, USA, 1985), pp. 1151–1154.
[3]F. Jacobsen, "Intensity measurements in the presence of moderate airflow," *Proc. Inter-Noise 94*, Yokohama, Japan (Noise Control Foundation, Poughkeepsie, NY, USA 1994), pp. 1737–1742.
[4]F. Jacobsen, "A note on measurement of sound intensity with windscreened probes," Appl. Acoust. **42**, 41–53 (1994).
[5]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier Publishing Company, New York, 1949), see Chap. I.
[6]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), see Sect. 6.2.
[7]J.-F. Allard and Y. Champoux, "New empirical equations for sound propagation in rigid frame fibrous materials," J. Acoust. Soc. Am. **91**, 3346–3353 (1992).
[8]D. K. Wilson, V. E. Ostashev, and S. L. Collier, "Time-domain equations for sound propagation in rigid-frame porous media," J. Acoust. Soc. Am. **116**, 1889–1892 (2004).
[9]A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "An advanced computational method for radiation and scattering of acoustic waves in three dimensions," J. Acoust. Soc. Am. **77**, 362–368 (1985).
[10]P. Juhl, "A numerical study of the coefficient matrix of the boundary element method near characteristic frequencies," J. Sound Vib. **175**, 39–50 (1993).
[11]F. Jacobsen, V. Cutanda, and P. M. Juhl, "A numerical and experimental investigation of the performance of sound intensity probes at high frequencies," J. Acoust. Soc. Am. **103**, 953–961 (1998).
[12]J. A. Mann III, J. Tichy, and A. J. Romano, "Instantaneous and time-averaged energy transfer in acoustic fields," J. Acoust. Soc. Am. **82**, 17–30 (1987).
[13]E. G. Williams, *Fourier Acoustics* (Academic Press, San Diego, 1999).
[14]M. Ren and F. Jacobsen, "A method of measuring the dynamic flow resistance and reactance of porous materials," Appl. Acoust. **39**, 265–276 (1993).
[15]F. Ingerslev, O. J. Pedersen, P. K. Møller, and J. Kristensen, "New rooms for acoustic measurements at the Danish Technical University," Acustica **19**, 185–199 (1967).
[16]IEC 61043 "Electroacoustics—Instruments for the measurement of sound intensity—Measurements with pairs of pressure sensing microphones" (International Electrotechnical Commission, Geneva, Switzerland, 1993).
[17]ANSI S1.9-1996 "Instruments for the Measurement of Sound Intensity" (American National Standards Institute, NY, USA, 1996).
[18]E. Frederiksen, "BCR-report: Free-field intensity sensitivity calibration and standing wave testing," (Brüel and Kjær, Nærum, Denmark, 1992).
[19]F. Jacobsen and E. S. Olsen, "Testing sound intensity probes in interference fields," Acustica **80**, 115–126 (1994).

A four particle velocity sensor device

D. R. Yntema, W. F. Druyvesteyn,[a] and M. Elwenspoek
Division of EEMCS, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 21 February 2005; revised 23 September 2005; accepted 21 November 2005)

In a sound field disturbance of pressure, particle velocity, density, temperature, and energy occur. In this paper acoustic disturbances in air are considered. In the majority of papers on acoustics only changes in the sound pressure are reported while in this paper results on the particle velocity are reported. Since particle velocity is a vector, while the pressure is a scalar, more information can be obtained when using a particle velocity sensor instead of a pressure sensor (microphone). Four particle velocity sensors are combined to one (small) device. In a reverberant room the four autospectra and the six cross spectra are determined. Interpretation of the measured results gives information of the free field (sound field without a contribution of reflections) as well as of the reverberant field. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151797]

PACS number(s): 43.58.Fm, 43.55.Mc [AJZi] Pages: 943–951

I. INTRODUCTION

In an acoustic disturbance in air, which is considered in this paper, (small) variations of the pressure p, the particle velocity v, the density ρ, or the temperature T, occur. Sensors for measuring the particle velocity in water are widely known, see Ref. 1. In air, often two closely spaced p sensors are used, from which the gradient of the pressure is estimated, which via the force equation is directly related to the particle velocity. Also a particle velocity sensor based on ultrasonic transduction has been proposed.[2] Two parallel ultrasonic beams are launched in opposite directions. The time difference in traveling time of the ultrasonic sound waves is proportional to the particle velocity. Nowadays a v sensor, the Microflown or μ flown,[3] is available and the particle velocity can be measured directly. Various experiments[3] have been performed to prove that with this sensor indeed the particle velocity is measured. Recently,[4] a direct comparison has been made by measuring the particle velocity at the same position with the method of Laser Doppler Anemometry (LDA), with the μ flown, a good agreement was found.

II. ONE V SENSOR

As a sensor for the particle velocity we used the "Microflown or μ flown."[3] This sensor consists basically of two thin wires, parallel to each other, which are heated up to about 200–300 °C by an electrical current; see Fig. 1. The temperature profile around the wires is influenced by the particle velocity of an acoustic disturbance, resulting in a small resistance change of the wires. In fact, the sensor detects the particle velocity component in a direction perpendicular to the length of the wires and in the plane of the wires. Define a unit vector $\boldsymbol{\mu}$, perpendicular to the length of the wires and in the plane of the wires, then the electrical signal of the sensor is proportional to the inner scalar product $\boldsymbol{\mu}.\mathbf{v}$. Throughout this paper we denote this unit vector as the sensitivity unit vector of the sensor. We often take the proportionality constant equal to one and write for the sensor signal $s(t)$: $s(t)=\boldsymbol{\mu}\cdot\mathbf{v}(t)$. The directional characteristic of this v sensor is independent of the frequency and follows from this relation as: $\boldsymbol{\mu}\cdot\mathbf{v}=|\mathbf{v}|\cdot\cos(\alpha_{\boldsymbol{\mu}\mathbf{v}})$, where $\alpha_{\mu v}$ is the angle between $\mathbf{v}$ and $\boldsymbol{\mu}$. A detailed model is presented in Ref. 5.

Since the measurement of the particle velocity is directly derived from the measurement of the electrical resistance of the sensor, the "Johnson" or resistance noise forms a lower limit of the noise of the microflown: $\langle V^2_{\text{noise}}\rangle_{f,f+\Delta f}=(4\ \text{kTR})\Delta f$, where k is the Boltzmann-constant, R the resistance, T the absolute temperature, and Δf the frequency interval or bandwidth. The noise can be higher than this lower limit, if a less optimal amplifier is used. Additionally the $1/f$ noise[6] is higher at low frequency than the resistance noise. In our experiments,[6] the noise level above about 600 Hz was indeed practically equal to the Johnson noise. At lower frequencies the $1/f$ noise dominates, see Fig. 2 and Ref. 6, the theoretical white noise level of a similar resistor is -165 dBV/$\sqrt{\text{Hz}}$.

A direct comparison of the noise or signal-to-noise ratio (SNR) between a v sensor and a p sensor can be done via the concept of the selfnoise. The selfnoise of a sensor is the ratio of the generated noise in volt and the sensitivity in volt/(m/sec) for a v sensor or volt/(Pa) for a p sensor; the obtained values represent the acoustical selfnoise of the device. Using the far field relation $v=p/\rho c$, with ρ the density of the air and c the sound velocity, the selfnoise times ρc of the v sensor can be compared with the selfnoise of the p sensor.

The state of the art is that at low frequencies the selfnoise times ρc of the v sensor is lower than the selfnoise of available p sensors, while at higher frequencies the selfnoise of the p sensor is lower.[7]

Consider now the sensor signal. For stationary signals the rms value, or the autospectrum, is measured, the latter being: $A=(1/T)\int_0^T s(t)s^*(t)dt$.

Suppose a sound source is present at the origin of a Cartesian coordinate system emitting plane waves, propagating in the x direction and a v sensor is placed on the x axis at a distance much larger than the wavelength and dimensions of the source (far field situation). When the sensitivity unit

[a]Electronic mail: w.f.druyvesteyn@utwente.nl

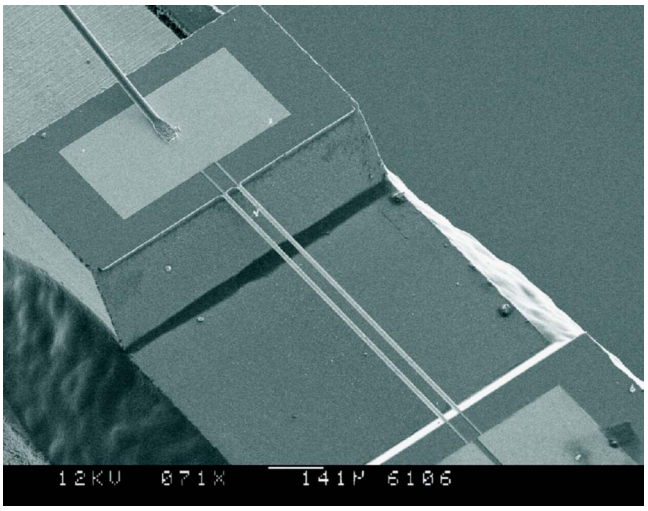


FIG. 1. The microflown sensor, SEM photo. The wires are both heating and sensing elements.

vector is along the x axis, $\boldsymbol{\mu}=[1,0,0]$ the free field and a contribution of the reverberant field will be measured. If however, the sensor is rotated such that $\boldsymbol{\mu}=[0,1,0]$ or $\boldsymbol{\mu}=[0,0,1]$ no contribution of the free field will be measured and only (a part) of the reverberant field is detected. In a pure diffuse reverberant sound field, where the distances from the mirror sources to the sensor are much larger than the wavelength and much larger than the dimensions of the sources, the relation between the rms value of the pressure and particle velocity is given by: $v_{\mathrm{rev}}^2=p_{\mathrm{rev}}^2/(\rho c)^2$. In a purely diffuse reverberant sound field the three components v_x^2, v_y^2, and v_z^2 are equal thus $v_x^2=v_{\mathrm{rev}}^2/3$ or $v_x^2=(1/3)p_{\mathrm{rev}}^2/(\rho c)^2$; this relation will be discussed in more detail in Sec. III.

When the v sensor is used for direct recording experiments with a sound source at the x axis it is interesting to listen to the cases where $\boldsymbol{\mu}=[1,0,0]$ and $\boldsymbol{\mu}=[0,1,0]$. For the former case the direct and reverberant field is heard, while for the latter case one listens to only the reverberant field. It is also interesting to listen to the reverberant field in the room at different positions.

III. TWO SENSORS

In this section the response of two v sensors for stationary signals is discussed; the orientation of the two v sensors is given by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Three spectra can now be determined, two autospectra and one cross spectrum, the latter being

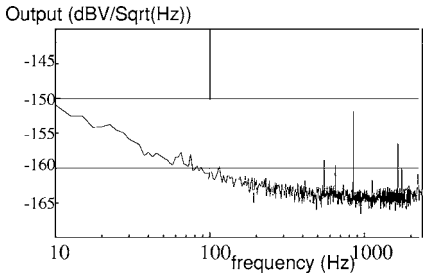


FIG. 2. Measured noise spectrum of a microflown.

$$C_{ij}=(1/T)\int_0^T s_i(t)s_j^*(t)dt. \tag{1}$$

The two autospectra A_i and A_j are found from Eq. (1) by taking $j=i$. Consider first the noise level of these signals. In most cases the noise level in the cross spectrum is lower than in the separate autospectra. The reason for this is that in the cross spectrum the product of the noise of sensor i and the noise of sensor j vanishes if the noise signals are uncorrelated and if the integration time T is taken long (the same holds for the product of the signal times the noise). Gordienko *et al.*[8] reported already in 1993 a reduction of the ocean noise field in the cross correlation of two sensors. Also, Shchurov[9] reported an increase in the signal-to-noise ratio (SNR) when an intensity probe is used. For isotropic (diffusive) noise even an increase in the SNR of 20–30 dB was found. For the case of two microflowns a reduction of the noise, down to 30 dB for frequencies above 600 Hz, has been reported.[6] The reason for this enormous reduction is that for frequencies above 600 Hz the dominant noise is the resistance noise (or Johnson noise). One expects that the resistance noise from one wire (resistance) powered by one battery is uncorrelated with the noise from another wire (resistance) powered by another battery. For lower frequencies $1/f$ noise appears, resulting in a much smaller reduction of the noise in the cross correlation.[6] The reason for this is that the decrease in the noise of the cross correlation spectrum is most effective for flat, broadband, frequency spectra. If a $1/f$ shaped component is present, there is a correlation between subsequent sample points x_i and x_j, while in a flat spectrum there is no correlation between points x_i and x_j $(i\neq j)$ in the series of data points.[6] Second another remarkable quality of the signal is that the influence of the reverberant sound field on the cross spectrum on one hand and the autospectra on the other hand is completely different. For the cross spectra the contribution of the reverberant field can even be zero, if $\boldsymbol{\mu}_1\cdot\boldsymbol{\mu}_2=0$, as will be explained below.

A simple two-dimensional case is worked out first. Two v sensors at the origin of the x-y plane are present, with $\boldsymbol{\mu}_1=[1,0]$ and $\boldsymbol{\mu}_2=[0,1]$. The two-dimensional purely diffuse reverberant sound field is modeled as four mirror sources $M_{1..4}$ on a circle making an angle β, $\pi-\beta$, $\pi+\beta$, and $2\pi-\beta$ with the x axis. Consider the contribution to the product of $s_1(t)$ and $s_2(t)$ when for each mirror source the velocity is directed inward or outward. For source M_1 the product of $s_1(t)s_2(t)$ will be positive [$s_1(t)$ and $s_2(t)$ are both negative for an inward velocity, since $\boldsymbol{\mu}_1=[1,0]$, $v_x<0$ and $\boldsymbol{\mu}_2=[0,1]$, $v_y<0$; $s_1(t)$ and $s_2(t)$ are both positive for an outward velocity, since $\boldsymbol{\mu}_1=[1,0]$, $v_x>0$ and $\boldsymbol{\mu}_2=[0,1]$, $v_y>0$. So the product $s_1(t)s_2(t)$ is positive]. For source M_2 the product $s_1(t)s_2(t)$ is negative [$s_1(t)>0$, $s_2(t)<0$ for an inward velocity and $s_1(t)<0$, $s_2(t)>0$ for an outward velocity], for M_3 the product is positive and for source M_4 again negative. So for all mirror sources, equally distributed over the circle and a long integration time, the total contribution will be zero.

Now consider the more general situation with a three-dimensional pure diffuse reverberant sound field with two v sensors in the origin with vectors $\boldsymbol{\mu}_1=[\mu_{1x},\mu_{1y},\mu_{1z}]$ and

$\boldsymbol{\mu}_2=[\mu_{2x},\mu_{2y},\mu_{2z}]$. Define first v^2_{rev}. Suppose a hypothetical omnidirectional v sensor is in the origin [this is certainly not a μ flown, which has a $\cos(\alpha_{\boldsymbol{\mu}\mathbf{v}})$ dependence, see Sec. I]. The mirror sources are equally distributed over a sphere. An arbitrary mirror source makes an angle α with the z axis and an angle β in the x-y plane, thus $v_x \propto \sin(\alpha)\cos(\beta)$, $v_y \propto \sin(\alpha)\sin(\beta)$, and $v_z \propto \cos(\alpha)$. Take the v_{rms} value from the sources within a solid angle $d\Omega=\sin(\alpha)d\alpha d\beta$, as v_m. The signal of the hypothetical omnidirectional sensor is proportional to $v^2_{\text{rev}}=\int v^2_m d\Omega=\int_0^{\pi}\sin(\alpha)d\alpha\int_0^{2\pi}d\beta\cdot v^2_m=4\pi v^2_m$. In the far field approximation $v^2_{\text{rev}}=p^2_{\text{rev}}/(\rho c)^2$.

The signal of the v sensor with $\boldsymbol{\mu}_1$ is $s_1=\boldsymbol{\mu}_1\cdot\mathbf{v}_\mathbf{m}=\mathrm{v_m}\{\mu_{1x}\cdot\sin(\alpha)\cos(\beta)+\mu_{1y}\cdot\sin(\alpha)\sin(\beta)+\mu_{1z}\cdot\cos(\alpha)\}$, thus the cross spectrum from the two v sensors with $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ becomes

$$C_{12}=\int_0^{\pi}\sin(\alpha)d\alpha\int_0^{2\pi}d\beta(\boldsymbol{\mu}_1\cdot\mathbf{v}_\mathbf{m})(\boldsymbol{\mu}_2\cdot\mathbf{v}_\mathbf{m})=v^2_{\text{rev}}(\boldsymbol{\mu}_1\cdot\boldsymbol{\mu}_2)/3. \tag{2}$$

Equation (2) follows from simple straightforward algebra and is written out in Appendix A.

Equation (2) shows that if $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are orthogonal to each other, there is no contribution of the reverberant sound field to the cross spectrum, and only the direct sound field of the source is measured. This is similar to the measurement of the real part of the sound intensity, for which also the reverberant sound field vanishes.[10] If $\boldsymbol{\mu}_1=\boldsymbol{\mu}_2$ (two v sensors in the same direction, in order to obtain a low noise level, as compared to the noise level in the autospectrum of one v sensor) the contribution of the reverberant field is $v^2_{\text{rev}}/3$. In Sec. IV the case of $\boldsymbol{\mu}_1=[1,1,0]/\sqrt{2}$ and $\boldsymbol{\mu}_2=[0,1,1]/\sqrt{2}$ is discussed; for this configuration the contribution of the reverberant field in the cross spectrum will be $v^2_{\text{rev}}/6$.

Thus when using two orthogonal oriented v sensors it is possible in a reverberant room to determine the direct sound field ("free field measurements") of a source, as well as the reverberant sound field in the room.

IV. MORE V SENSORS

When using two v sensors it is possible that with a sound source in a room no direct sound is detected in A_1, A_2, and C_{12}. For example, if the sound source is at [0,0,0] and the sensors at [1,1,0] with $\boldsymbol{\mu}_1=[-1,1,0]$ and $\boldsymbol{\mu}_2=[0,0,1]$ no direct sound is detected: $A_1=A_2=v^2_{\text{rev}}/3$ and $C_{12}=0$. It seems logical then to use three v sensors with $\boldsymbol{\mu}_1=[1,0,0]$, $\boldsymbol{\mu}_2=[0,1,0]$, and $\boldsymbol{\mu}_3=[0,0,1]$ and to use the three low-noise cross spectra for the source localization. However, for that case it may occur that all three cross spectra are zero; e.g., source at [0,0,0] and sensors at x axis (or y axis or z axis).

In order to anticipate this, four v sensors should be used. In the device discussed below we used: $\boldsymbol{\mu}_1=[1,0,1]/\sqrt{2}$, $\boldsymbol{\mu}_2=[0,1,1]/\sqrt{2}$, $\boldsymbol{\mu}_3=[-1,0,1]/\sqrt{2}$, and $\boldsymbol{\mu}_4=[0,-1,1]/\sqrt{2}$. As measured spectra we now have four autospectra and six cross spectra. The contribution of the free field to the time averaged auto- and cross-spectra can be found as follows. Write the free field particle velocity vector and its time dependence as $\mathbf{v}(v_x,v_y,v_z)a(t)$, where the vector $\mathbf{v}(v_x,v_y,v_z)$, with (the real) components v_x, v_y, and v_z represents the directional vector of the free field particle velocity and $a(t)$ the time dependence of the source signal. Write for the signal of sensor A, divided by the sensor sensitivity [in V/(m/sec)], the symbol $S_A(t)$, then $s_A(t)=(\boldsymbol{\mu}_A\cdot\mathbf{v})a(t)=(\mu_{xA}\cdot v_x+\mu_{yA}\cdot v_y+\mu_{zA}\cdot v_z)a(t)$; $\boldsymbol{\mu}_A=[\mu_{xA},\mu_{yA},\mu_{zA}]$ is the sensitivity unit vector of sensor A. The time averaged autospectrum of sensors A is thus equal to $\overline{(\mu_{xA}\cdot v_x+\mu_{yA}\cdot v_y+\mu_{zA}\cdot v_z)^2a(t)a^*(t)}$. Similarly the time averaged cross spectrum of the signals of sensors A and B, the latter with sensitivity unit vector $\boldsymbol{\mu}_B=[\mu_{xB},\mu_{yB},\mu_{zB}]$ is equal to

$$\overline{(\boldsymbol{\mu}_A\cdot\mathbf{v})(\boldsymbol{\mu}_B\cdot\mathbf{v})a(t)a^*(t)}=\overline{(\mu_{xA}\cdot v_x+\mu_{yA}\cdot v_y+\mu_{zA}\cdot v_z)(\mu_{xB}\cdot v_x+\mu_{yB}\cdot v_y+\mu_{zB}\cdot v_z)a(t)a^*(t)}.$$

In the auto- and cross-spectrum the same term $a(t)a^*(t)$ appears, and in a comparison between these spectra this term can thus be omitted. For the four v sensors, as defined above, the four autospectra and the six cross spectra are

$$A_1=v_x^2/2+v_z^2/2+v_xv_z+v^2_{\text{rev}}/3, \tag{3}$$

$$A_2=v_y^2/2+v_z^2/2+v_yv_z+v^2_{\text{rev}}/3, \tag{4}$$

$$A_3=v_x^2/2+v_z^2/2-v_xv_z+v^2_{\text{rev}}/3, \tag{5}$$

$$A_4=v_y^2/2+v_z^2/2-v_yv_z+v^2_{\text{rev}}/3, \tag{6}$$

$$C_{12}=v_z^2/2+v_xv_y/2+v_xv_z/2+v_yv_z/2+v^2_{\text{rev}}/6, \tag{7}$$

$$C_{13}=-v_x^2/2+v_z^2/2, \tag{8}$$

$$C_{14}=v_z^2/2-v_xv_y/2+v_xv_z/2-v_yv_z/2+v^2_{\text{rev}}/6, \tag{9}$$

$$C_{23}=v_z^2/2-v_xv_y/2-v_xv_z/2+v_yv_z/2+v^2_{\text{rev}}/6, \tag{10}$$

$$C_{24}=-v_y^2/2+v_z^2/2, \tag{11}$$

$$C_{34}=v_z^2/2+v_xv_y/2-v_xv_z/2-v_yv_z/2+v^2_{\text{rev}}/6. \tag{12}$$

In an experiment one thus determines ten spectra and there are four unknown quantities, v_x, v_y, v_z, and v_{rev}. As an approximation for the best solution of Eqs. (3)–(12) the following algebraic procedure was used. From Eqs. (3)–(12) the products of $v_i\cdot v_j$ are estimated

$$v_x\cdot v_y=\{C_{12}-C_{14}-C_{23}+C_{34}\}/2=C_{xy}, \tag{13}$$

$$v_x \cdot v_z = \{A_1 - A_3 + C_{12} + C_{14} - C_{23} - C_{34}\}/4 = C_{xz}, \quad (14)$$

$$v_y \cdot v_z = \{A_2 - A_4 + C_{12} - C_{14} + C_{23} - C_{34}\}/4 = C_{yz}. \quad (15)$$

Use as variables $u=v_x^2$, $v=v_y^2$, and $w=v_z^2$, then one obtains the six following equations:

$$u - v = A_1 - A_2 + A_3 - A_4 = A_0, \quad (16)$$

$$-u + w = 2C_{13}, \quad (17)$$

$$-v + w = 2C_{24}, \quad (18)$$

$$u = (v_x \cdot v_y)(v_x \cdot v_z)/(v_y \cdot v_z) = C_{xy} \cdot C_{xz}/C_{yz} = C_u, \quad (19)$$

$$v = (v_x \cdot v_y)(v_y \cdot v_z)/(v_x \cdot v_z) = C_{xy} \cdot C_{yz}/C_{xz} = C_v, \quad (20)$$

$$w = (v_x \cdot v_z)(v_y \cdot v_z)/(v_x \cdot v_y) = C_{yz} \cdot C_{xz}/C_{xy} = C_w. \quad (21)$$

These six equations with three unknown variables are solved using the least mean squares method; the expression

$$(u - v - A_0)^2 + (-u + w - 2C_{13})^2 + (-v + w - 2C_{24})^2 + (u - C_u)^2 + (v - C_v)^2 + (w - C_w)^2, \quad (22)$$

should have a minimum value. Differentiating with respect to the three variables gives

$$\partial/\partial u = 0: \quad 3u - v - w - A_0 + 2C_{13} - C_u = 0, \quad (23)$$

$$\partial/\partial v = 0: \quad -u + 3v - w + A_0 + 2C_{24} - C_v = 0, \quad (24)$$

$$\partial/\partial w = 0: \quad -u - v + 3w - 2C_{13} - 2C_{24} - C_w = 0. \quad (25)$$

Solving these equations gives as the best solutions

$$4u = 4v_x^2 = A_0 - 2C_{13} + 2C_u + C_v + C_w, \quad (26)$$

$$4v = 4v_y^2 = -A_0 - 2C_{24} + C_u + 2C_v + C_w, \quad (27)$$

$$4w = 4v_z^2 = 2C_{13} + 2C_{24} + C_u + C_v + 2C_w, \quad (28)$$

from which the free field particle velocity vector and the reverberant field can be solved.

If the SNR (signal-to-noise ratio) in A_1–A_4 is low, but high in C_{ij} then the autospectra should not be used in Eqs. (14) and (15).

An alternative and practical method to obtain the best solution of Eqs. (3)–(12) is to use a simple straightforward numerical procedure (on a PC) to find the minimum value of the expression

$$(v_x^2/2 + v_z^2/2 + v_x v_z + v_{\text{rev}}^2/3 - A_1)^2 + (v_y^2/2 + v_z^2/2 + v_y v_z + v_{\text{rev}}^2/3 - A_2)^2 + (v_x^2/2 + v_z^2/2 - v_x v_z + v_{\text{rev}}^2/3 - A_3)^2 + (v_y^2/2 + v_z^2/2 - v_y v_z + v_{\text{rev}}^2/3 - A_4)^2 + (v_z^2/2 + v_x v_y/2 + v_x v_z/2 + v_y v_z/2 + v_{\text{rev}}^2/6 - C_{12})^2 + (-v_x^2/2 + v_z^2/2 - C_{13})^2 + (v_z^2/2 - v_x v_y/2 + v_x v_z/2 - v_y v_z/2 + v_{\text{rev}}^2/6 - C_{14})^2 + (v_z^2/2 - v_x v_y/2 - v_x v_z/2 + v_y v_z/2 + v_{\text{rev}}^2/6 - C_{23})^2 + (-v_y^2/2 + v_z^2/2 - C_{24})^2 + (v_z^2/2 + v_x v_y/2 - v_x v_z/2 - v_y v_z/2 + v_{\text{rev}}^2/6 - C_{34})^2.$$

The values of v_x, v_y, and v_z, as found from Eqs. (25)–(27) and $v_{\text{rev}}^2 \geqslant 0$ can be used as starting values in this numerical procedure.

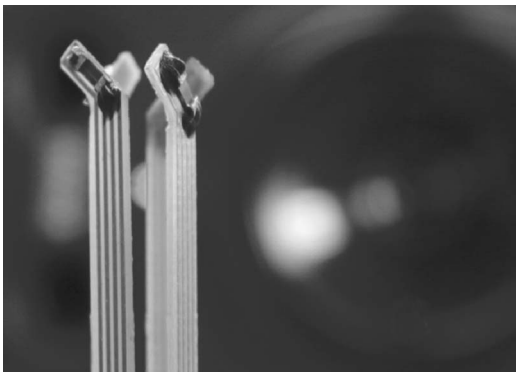


FIG. 3. The four sensor device, microflown elements bonded to an epoxy carrier.

V. THE FOUR SENSOR DEVICE

A photograph of the four v sensor device is shown in Fig. 3. Four small printed circuit boards, each with one μ flown on it are combined. The orientations of the four v sensors are $\boldsymbol{\mu}_1=[1,0,1]/\sqrt{2}$, $\boldsymbol{\mu}_2=[0,1,1]/\sqrt{2}$, $\boldsymbol{\mu}_3=[-1,0,1]/\sqrt{2}$, and $\boldsymbol{\mu}_4=[0,-1,1]/\sqrt{2}$. For the configuration, as shown in Fig. 3 the directional characteristics of each sensor corresponds to $\cos(\alpha_{\boldsymbol{\mu}\mathbf{v}})$, see Fig. 4. It is necessary that the encapsulation is not done by combining the four printed circuit boards on a fixed bar, since this will lead to (frequency dependent) deviations of the directional characteristics as given in Sec. I as $\cos(\alpha_{\boldsymbol{\mu}\mathbf{v}})$. The distance between the four sensors is only 5 mm. The main electronics are encapsulated in the holder and a seven-pin connector is used for the power supply and connection to a sound card or analyzer.

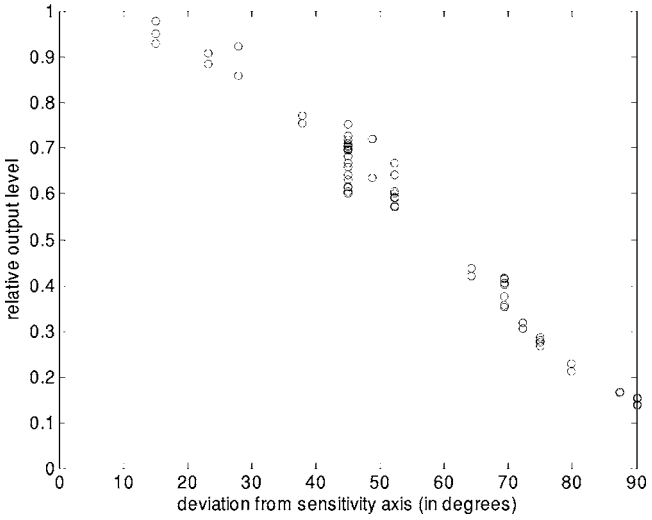


FIG. 4. Directionality measurement of a microflown element, relative sensitivity against deviation from maximum sensitivity direction.

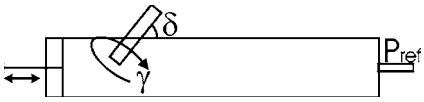


FIG. 5. Standing wave tube, a calibration device.

The calibration of the v sensors has been described extensively in a number of papers.[3,4,6] There are different methods to obtain a calibrated v sensor; three methods are mentioned below.

1. A calibrated p sensor and the v sensor in a standing wave tube and using the relation between the position dependence of the pressure and particle velocity.
2. A calibrated p sensor and a v sensor in a free field and far field situation, where $v=p/(\rho c)$.
3. Measure the particle velocity using the Laser Doppler Anemometry (LDA) method,[4] and compare it with the signal from the v sensor. The LDA is an independent method to measure the particle velocity.[4]

A practical method is to use a one-dimensional standing wave tube; a schematic drawing is shown in Fig. 5. The method to obtain the sensitivity of the v sensor as a function of frequency from measurements in a standing wave tube (SWT), has been described by Raangs *et al.*[4] The microphone is positioned at the end of the tube ($x=1$), the v sensor at a position x, where $v(x)$ is measured. A relation between $v(x)$ and $p(x=1)$ is used, where the damping is taken into account in a standing wave ratio (SWR)

$$\frac{v(x)}{p(l)} = \frac{1}{\rho_0 c}\left[\frac{1}{SWR}\cos\{k(\ell - x) - \vartheta/2\} + j\sin\{k(l-x) - \vartheta/2\}\right], \tag{29}$$

with ϑ a phase angle.

In the case where the damping is neglected, a simple relation remains. In our experiments the damping was quite small, a SWR of about 20–25 dB was measured. The measured data are corrected using Eq. (29) and are then fitted with a good approximation of a Microflown[11] output $= \frac{LFS}{\sqrt{(1+f^2/f^2_{\text{heat cap}})(1+f^2/f^2_{\text{diff}})}}$, with LFS the low frequency sensitivity, f_{diff} the diffusion corner frequency, and $f_{\text{heat cap}}$ the heat capacity corner frequency. It should be noticed that for interpretation of our experiments the differences between the sensitivities are of importance, and not the absolute sensitivity; systematic errors in the sensitivities of the four sensors do not play a role.

Simple relations can be derived for the position dependence and the relation between pressure and particle velocity. A number of p and v sensors can therefore be placed at different positions in the standing wave tube. In Fig. 5 the (not calibrated) v sensor is positioned under an angle δ with the standing wave tube, together with calibrated, reference, p and v sensors.

The four v sensors (and the x-y axes) can be rotated around the probe holder, which is defined as the z axis; the rotation angle $\gamma=0$ corresponds with the drawing in Fig. 5, i.e., the x axis in the plane of the drawing and the y axis perpendicular to it. Due to the rotation around the z axis the particle velocity in the standing wave tube is written in this Cartesian coordinate system as: $\mathbf{v} \propto [\cos(\delta)\cos(\gamma), \cos(\delta)\sin(\gamma), \sin(\delta)]$; the four unit vectors of the v sensors remain constant and equal to the values given above. The directional characteristics of for instance v sensor one thus become $\mathbf{v}\cdot\mathbf{n}_1 \propto \cos(\delta)\cos(\gamma)+\sin(\delta)$ and of v sensor two $\mathbf{v}\cdot\mathbf{n}_2 \propto \cos(\delta)\sin(\gamma)+\sin(\delta)$.

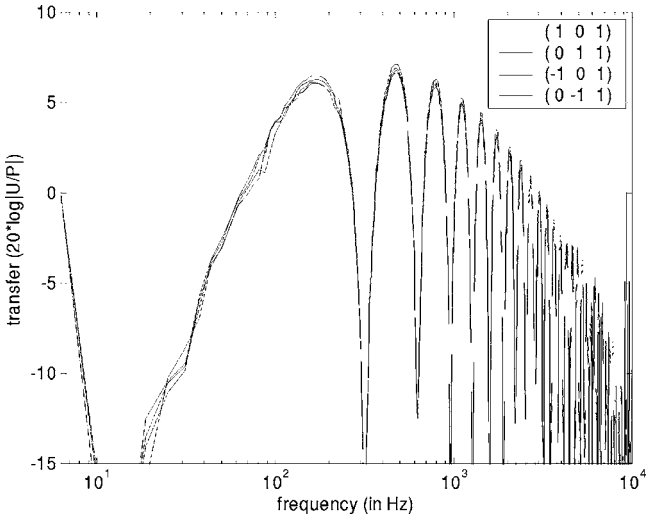


FIG. 6. Transfer function between microflown and reference microphone (14 mV/Pa) in a standing wave tube. Calibration values can be obtained at the maxima. (Differences in sensitivity are approximately 1/2 dB.)

In Fig. 6 the measured output in the $\boldsymbol{\mu}_i$ direction as a function of frequency is shown for the four v sensors. The curves show that the differences between the four v sensors are quite small.

VI. EXPERIMENTS IN A REVERBERANT ROOM

In a reverberant room with dimensions of about 9^*7^*4 m a loudspeaker box with dimensions $0.4^*0.2^*0.2$ m was used as a sound source. The four v-sensor device was placed at various locations in the room, in such a way that the inclination (angle δ with the probe holder) and rotation angle (angle γ in a plane perpendicular to the probe holder) were varied. The source-sensor distance was also varied in order to create a situation where $v^2_{\text{direct}}/v^2_{\text{reverberant}} < 1$ and $v^2_{\text{direct}}/v^2_{\text{reverberant}} > 1$. When the inclination angle was taken as δ, the rotation angle as γ and the sound source-sensor distance as r, the response was measured for $r=0.5$, 1, and 2 m, δ as well as $\gamma=0$, 30, 60, and 90 deg; so in total 48 responses. As the signal to the loudspeaker white noise was used; the same (white noise) time signal was used in the 48 experiments. The auto- and cross-spectra were averaged over 4 s.

A four input channel device (product name Siglab) and a sound card in a PC were used as input for calculating the auto- and cross-spectra.

TABLE I. Results for one meter distance to the source, the bold numbers are actual angles, the measured angles are given in the table.

Incl, $\delta\downarrow$ \Rot, γ	**0**	**30**	**60**	**90**
0	−0.79 \ 1.2	1.47 \ 31	0.93 \ 61.1	0.74 \ 89.6
30	29.1 \ 1.1	29.0 \ 28.3	29.4 \ 57.8	28.2 \ 86.6
60	57.4 \ −1.9	57.0 \ 29.8	57.6 \ 59.5	56.1 \ 91.1
90	87.7 \ (17)	86.0 \ (−7)	86.7 \ (−9)	85.9 \ (−30)

TABLE II. Measurement and errors, β is the deviation in degrees averaged over 16 measurements.

r (m)	$v_{\mathrm{dir}}^2/v_{\mathrm{rev}}^2$	$\cos(\beta)$	β (deg)	$\sin(\beta)$	σ	β (deg)
0.5	13.2	0.9984	3.3	0.0513	0.025	3
1	3.15	0.9989	2.7	0.0435	0.017	2.5
2	0.72	0.9900	8	0.1322	0.049	7.5

VII. RESULTS

In acoustic measurements it may well be that one is primarily interested in the free field properties of a sound source; on the other hand there are situations where one is primarily interested in the properties of the reverberant sound field. An example of the first case is that the one wants to know the radiated power (noise) of a machine while it is placed in a reverberant room. An example of the second case is that one is interested in the quality of a concert hall where the properties of the reverberant field are of interest. When using a p sensor it is quite difficult or even impossible to measure the free and the reverberant field separately. When the distance r, between sound source and sensor is larger than the reverberation distance r_r, of the room it is difficult to measure accurately the free field on the other hand, when the distance is small it is difficult to determine the properties of the reverberant field (the reverberation distance r_r is defined as the distance from the source where the rms value of the free field is equal to the rms value of the reverberant field). However with the 4-v-sensor device a number of possibilities appear. In Sec. VII A, the emphasis is on the determination of the free field, in Sec. VII B, the determination of the reverberant field is discussed. In the next two sections it will be shown that even in the case of $r>r_r$ it is possible to determine the free field, while for the case of $r<r_r$ the properties of the reverberant field can be determined.

A. Free field, localization of the sound source

The free or direct sound field is calculated from the measured ten auto- and cross-spectra as was explained in Sec. III. From the ratios of v_x/v_y, v_x/v_z, and v_y/v_z, of the direct field the direction of the sound source seen from the sensor is calculated and compared with the theoretical value as deduced from the experimental set up.

In Table I results for the case $r=1$ m are given; the bold numbers refer to the angles in degrees of the experimental setup, the "theoretical angles." The other numbers for the angles in degrees are the results of the calculations from the experimental results. In, e.g., 29.1 \ 1.1 the value of 29.1 refers to the value of δ, and 1.1. to the value of γ.

The (numerical) calculated values for v_x and v_y for the case $\delta=90$ have no meaning and are placed between brackets. When $\delta=90$, $v_x=v_y=0$ and a value of arctan (v_x/v_y) is undefined. In practice such a situation is recognized when the solutions for the components of the vector v are such that $v_x \ll v_z$ and $v_y \ll v_z$ while v_x and v_y are of the same order.

Of course the numerical procedure gives values for v_x and v_y, but the ratio in arctan (v_x/v_y) has no meaning here.

The average reverberant sound field v_{rev}^2 is also calculated from these experimental values. In the second column of Table II an averaged value of the ratio direct/reverberant particle velocity, $v_{\mathrm{dir}}^2/v_{\mathrm{rev}}^2$ is given. For a more detailed comparison between experimental and theoretical directions the inner, scalar, vector product of the theoretical particle velocity vector, taken as unit vector and denoted as $\mathbf{v}_{\mathrm{th}}$, and the experimental particle velocity vector, denoted as $\mathbf{v}_{\mathrm{exp}}$, is calculated. The deviation between the directions of $\mathbf{v}_{\mathrm{th}}$ and $\mathbf{v}_{\mathrm{exp}}$, denoted as β is then found from: $\cos(\beta)=|\mathbf{v}_{\mathrm{th}}\cdot\mathbf{v}_{\mathrm{exp}}|/|\mathbf{v}_{\mathrm{exp}}|$. This deviation angle β can also be found from the inner vector product as: $\sin(\beta)=|\mathbf{v}_{\mathrm{th}}\boldsymbol{x}\mathbf{v}_{\mathrm{exp}}|/|\mathbf{v}_{\mathrm{exp}}|$, where the symbol $\boldsymbol{x}$ is used for the vector product. In terms of δ and γ the theoretical particle velocity vector is written as $\mathbf{v}_{\mathrm{th}}=\mathrm{i}[\cos(\delta)\cos(\gamma)]+\mathrm{j}[\cos(\delta)\sin(\gamma)]+\mathrm{k}[\sin(\delta)]$. In the third and fifth columns of Table II, the averaged values of $\cos(\beta)$ and $\sin(\beta)$ are given, the averaging is done over the 16 different directions given by δ and γ. For the deviation expressed by $\sin(\beta)$ also a standard deviation σ is given, calculated from: $\sigma=\sqrt{\mathrm{var}}=\frac{1}{n}\sqrt{\Sigma(\mathrm{value}-\mathrm{mean})^2}$, with $n=16$. The table shows a fairly good similarity between experimental found direction of the source and the real direction. So it is thus possible to measure the direct sound field in a reverberant environment using this 4-v-sensor device. Even for the case where the sound source-sensor distance is larger than the reverberation distance the results are within 8 deg.

B. The reverberant field

The fact that in the autospectrum of a v sensor only $v_{\mathrm{rev}}^2/3$ is measured means that only a part of the reverberant field contributes to it. Thus by measuring a number of autospectra (or cross spectra) information about different parts of the reverberant field can be obtained. For example, if there are three v sensors, with $\boldsymbol{\mu}_1=[1,0,0]$, $\boldsymbol{\mu}_2=[0,1,0]$, and $\boldsymbol{\mu}_3=[0,0,1]$, then the signal of sensor one will contain the reverberant field around the x axis, sensor two the field around the y axis, and sensor three the field around the z axis.

When the free field is known, it is in principle possible to calculate the reverberant field from the difference of the autospectrum and the free field. However, this will not be a good method when $r<r_r$ since the reverberant field is smaller than the free field; consider, e.g., the situation with $r=0.5$ m in Table II, where $v_{\mathrm{dir}}^2/v_{\mathrm{rev}}^2=13.2$. It seems better to calculate it from experimental values, which, in the case of no reflections, should be zero. An example is the expression $A_i\cdot A_j-C_{ij}^2$. In the case of no reflections $A_i=(\mathbf{v}\cdot\boldsymbol{\mu}_i)^2$, $A_j=(\mathbf{v}\cdot\boldsymbol{\mu}_j)^2$, and $C_{ij}=(A_i\cdot A_j)=(\mathbf{v}\cdot\boldsymbol{\mu}_i)(\mathbf{v}\cdot\boldsymbol{\mu}_j)$ thus $A_i\cdot A_j-C_{ij}^2$ vanishes. In the case of reflections $A_i\cdot A_j-C_{ij}^2$ can be written as a quadratic expression of v_{rev}^2, which then can be solved;

TABLE III. Reverberant field v^2_{rev}, calculated from two sensors at the time, values were normalized with respect to v^2_{dir} at $r=1$ m.

$r \backslash \gamma \backslash \delta$	v^2_{rev} (1,2)	v^2_{rev} (1,3)	v^2_{rev} (1,4)	v^2_{rev} (2,3)	v^2_{rev} (2,4)	v^2_{rev} (3,4)
0.5\60\ 0	0.400	0.413	0.360	0.280	0.320	0.288
0.5\90\ 0	0.331	0.251	0.332	0.266	0.422	0.265
0.5\90\ 30	0.341	0.347	0.377	0.435	0.368	0.338
0.5\0\ 60	0.398	0.406	0.325	0.404	0.301	0.379
1\30\ 0	0.323	0.377	0.304	0.286	0.330	0.289
1\60\ 30	0.331	0.274	0.317	0.274	0.318	0.265

the simple and straightforward algebra for the cases $i=1$, $j=2$, and $i=1$ and $j=3$ is given in Appendix B. A reason for taking $A_i \cdot A_j - C^2_{ij}$ is that this quantity is not sensitive to deviations in the sensitivity or orientation vector $\boldsymbol{\mu}$ of the separate sensors, or to a (small) deviation in the orientation of the 4-v-sensor with respect to the desired experimental setup. However, the noise and SNR ratio can be different for A_i and C_{ij}.

For the 4-v-sensor device there are six independent values of $A_i \cdot A_j - C^2_{ij}$, thus six different contributions of the reflections to the total reverberant field.

Some examples of results for v^2_{rev} are given in Table III. In this table $r\backslash\gamma\backslash\delta=0.5\backslash60\backslash0$ refers to the experimental setup where the distance is 0.5 m, the inclination $\delta=0$ and the angle $\gamma=60$ deg; v^2_{rev} (1,2) is the result using $A_1 \cdot A_2 - C^2_{12}$. The values of v^2_{rev} in Table III have been normalized with respect to the average value of v^2_{dir} for $r=1$ m. This is done because a value of v^2_{rev} in $(m/s)^2$ does not give any information, unless the excitation used is known, together with the knowledge of a number quantities as, e.g., the loudspeaker efficiency, the radiation impedance, etc. A direct measure for comparison is the direct (or free) sound field v^2_{dir} (the choice of $r=1$ m is of course arbitrary).

So by measuring and calculating the parts of the reverberant field information can be obtained about the contributions from different directions to the total reverberant sound field, which in turn gives information about the acoustics of the room.

In this paper, with emphasis on the 4-v-sensor device, we will not discuss these aspects.

In general one is inclined to think that the average reverberant sound field is independent of the distance r. For the three distances $r=0.5$, 1, and 2 m the results are given in Table IV; in the second column the average value of v^2_{rev}, again normalized with respect to $v^2_{dir}(r=1$ m), in the third column the standard deviation σ. The values of v^2_{rev} for the different distances are not exactly equal; however their differences are within the standard deviation.

TABLE IV. Reverberant field measured, normalized with respect to v^2_{dir} ($r=1$ m), and standard deviation.

r (m)	v^2_{rev}	σ
0.5	0.358	0.066
1	0.310	0.041
2	0.299	0.074

VIII. CONCLUSIONS AND FUTURE PLANS

A four-particle velocity sensor has been described with which the free field and the reverberant sound field can be detected in a reverberant environment. The strength of these sound fields is not expressed in terms of the sound pressure but in terms of the particle velocity. Since the latter is a vector, while the pressure is a scalar, more information can be obtained when using a particle velocity sensor instead of a pressure sensor (microphone). Free field measurements are desired for determining the radiated noise of a source, when it is positioned in a reverberant environment. In (concert) hall acoustics a precise and detailed knowledge of the reflection pattern and reverberant sound field is desired. From measurements in a reverberant room the free field sound field, as well as the reverberant sound have been deduced. This has been done for a source-sensor distance as well as smaller and larger than the reverberation distance of the room. Also information about the directional characteristics of the reverberant field is obtained by comparing the various auto and cross correlations of the four sensors.

The separate four v sensors are quite small ($4^*1.5^*1.5$ mm) and are encapsulated close to each other, so that the sensing volume (the volume in which the three vector components of the particle velocity are measured) is small as well; about 5^*5^*5 mm.

Many separate v sensors were processed on a wafer in a thin film technology. They were cut, glued on a small printed circuit boards with electrical contact areas. The four printed circuit boards were encapsulated together to one device. The next step in the integration process is to integrate all sensors on one die (a piece of silicon), a photo of this developed

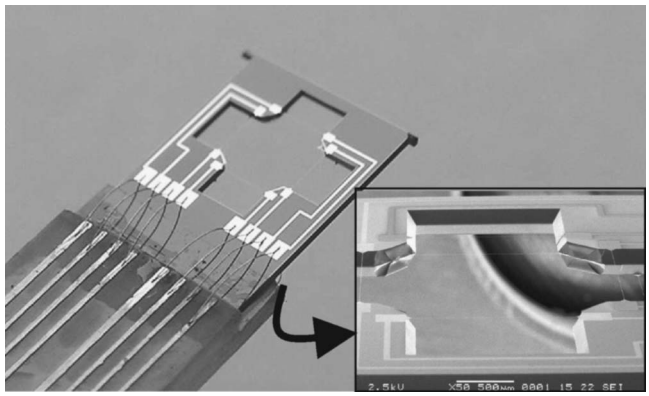


FIG. 7. Photo of the newly made device. The sensor wires can be seen inside the cross shape as thin lines.

sensor is shown in Fig. 7. The mounting of the sensor is much simpler (only one printed circuit board is needed with electrical contacts), the directions of the four unit vectors $\boldsymbol{\mu_i}$ are now precisely defined by a lithography process step, and also the equality of the four v sensors in one device is even better.

APPENDIX A

$$
\begin{aligned}
(\boldsymbol{\mu}_1 \cdot \mathbf{v}_m)(\boldsymbol{\mu}_2 \cdot \mathbf{v}_m) = v_m^2\{&\mu_{1x} \cdot \mu_{2x} \cdot \sin^2(\alpha) \cdot \cos^2(\beta) \\
&+ \mu_{1y} \cdot \mu_{2y} \cdot \sin^2(\alpha)\sin^2(\beta) \\
&+ \mu_{1z} \cdot \mu_{2z} \cdot \cos^2(\alpha)\} \\
&+ v_m^2[\mu_{1x} \cdot \mu_{2y} \cdot \sin^2(\alpha)\cos^2(\beta)\sin(\beta) \\
&+ \mu_{1x} \cdot \mu_{2z} \cdot \sin(\alpha)\cos(\alpha)\cos(\beta) \\
&+ \mu_{1y} \cdot \mu_{2x} \cdot \sin^2(\alpha)\cos^2(\beta)\sin(\beta) \\
&+ \mu_{1y} \cdot \mu_{2z} \cdot \sin(\alpha)\cos(\alpha)\sin(\beta) \\
&+ \mu_{1z} \cdot \mu_{2x} \cdot \sin(\alpha)\cos(\alpha)\cos(\beta) \\
&+ \mu_{1z} \cdot \mu_{2y} \cdot \sin(\alpha)\cos(\alpha)\sin(\beta)].
\end{aligned}
$$

The integration $\int_0^{\pi}\sin(\alpha)d\alpha \int_0^{2\pi}d\beta$ of the terms between [ ] results in a zero value (e.g., $\int_0^{2\pi}\cos(\beta)\sin(\beta)d\beta=0$, etc.). The integration $\int_0^{\pi}\sin(\alpha)d\alpha\int_0^{2\pi}d\beta$ of the terms between { } results in

$$
\begin{aligned}
&(4\pi/3)v_m^2\{\mu_{1x} \cdot \mu_{2x} + \mu_{1y} \cdot \mu_{2y} + \mu_{1z} \cdot \mu_{2z}\} \\
&\quad = v_{\mathrm{rev}}^2(\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2)/3
\end{aligned}
$$

(e.g., $v_m^2\int_0^{\pi}\sin(\alpha)d\alpha\int_0^{2\pi}\sin^2(\alpha)\cdot\cos^2(\beta)d\beta=v_m^2\pi\int_0^{\pi}\sin^3(\alpha)d\alpha = 4\pi v_m^2/3=v_{\mathrm{rev}}^2/3$ or $v_m^2\int_0^{\pi}\sin(\alpha)d\alpha\int_0^{2\pi}\cos^2(\alpha)d\beta = -v_m^2 2\pi\int_0^{\pi}\cos^2(\alpha)d\cos(\alpha)=4\pi v_m^2/3=v_{\mathrm{rev}}^2/3$).

Instead of writing out all the components of the arbitrary vectors $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ a faster derivation of Eq. (2) is possible. Since the integration should be performed over all mirror sources equally distributed over the sphere the result of the integration is independent of any rotation of the Cartesian coordinate system. Rotate then the coordinate system such that $\boldsymbol{\mu_1}=[0,0,1]$ and $\boldsymbol{\mu_2}=[0,\sin(\alpha_\mu),\cos(\alpha_\mu)]$. The angle between $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ is α_μ and the inner product $\boldsymbol{\mu_1}\cdot\boldsymbol{\mu_2}$ is equal to $\cos(\alpha_\mu)$. The integral equation for the contribution of the reverberant field to the cross spectrum becomes:

$$
\begin{aligned}
&v_m^2\int_0^{\pi}\sin(\alpha)\cdot d\alpha\cdot\int_0^{2\pi}d\beta\{[\cos(\alpha)] \\
&\times[\sin(\alpha_\mu)\sin(\alpha)\sin(\beta)+cos(\alpha_\mu)\cos(\alpha)]\} \\
&=2\pi v_m^2\int_0^{\pi}\sin(\alpha)\cos^2(\alpha)\cos(\alpha_\mu)d\alpha=(4\pi v_m^2/3)\cos(\alpha_\mu) \\
&=(v_{\mathrm{rev}}^2/3)(\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2).
\end{aligned}
$$

APPENDIX B

$$
\begin{aligned}
A_1 \cdot A_2 - (C_{12})^2 = \{&v_x^2/2 + v_z^2/2 + v_x \cdot v_z + v_{\mathrm{rev}}^2/3\}\{v_y^2/2 \\
&+ v_z^2/2 + v_y \cdot v_z + v_{\mathrm{rev}}^2/3\} \\
&- \{v_z^2/2 + v_x \cdot v_y/2 + v_x \cdot v_z/2 + v_y \cdot v_z/2 + v_{\mathrm{rev}}^2/6\}^2
\end{aligned}
$$

$$
\begin{aligned}
=&v_x^2 \cdot v_y^2/4 + v_x^2 \cdot v_z^2/4 + v_x^2 \cdot v_y \cdot v_z/2 + (v_{\mathrm{rev}}^2/3)(v_x^2/2 + v_z^2/2 \\
&+ v_x \cdot v_z + A_2) + v_y^2 \cdot v_z^2/4 + v_z^4/4
\end{aligned}
$$

$$
\begin{aligned}
&+ v_y \cdot v_z^3/2 + v_x \cdot v_y^2 \cdot v_z/2 + v_x \cdot v_z^3/2 + v_x \cdot v_y \cdot v_z^2 - \{v_z^4/4 \\
&+ v_x^2 \cdot v_y^2/4 + v_x^2 \cdot v_z^2/4 + v_y^2 \cdot v_z^2/4
\end{aligned}
$$

$$
\begin{aligned}
&+ (v_{\mathrm{rev}}^2/6)^2 + v_x \cdot v_y \cdot v_z^2/2 + v_x \cdot v_z^3/2 + v_y \cdot v_z^3/2 + (v_{\mathrm{rev}}^2/3) \\
&\times(v_z^2/2 + v_x \cdot v_y/2 + v_x \cdot v_z/2 + v_y \cdot v_z/2)
\end{aligned}
$$

$$
\begin{aligned}
&+ v_x^2 \cdot v_y \cdot v_z/2 + v_x \cdot v_y^2 \cdot v_z/2 + v_x \cdot v_y \cdot v_z^2/2 \\
&=(v_{\mathrm{rev}}^2/3)(A_1 + A_2 - v_{\mathrm{rev}}^2/3) - \{(v_{\mathrm{rev}}^2/6)^2 + (v_{\mathrm{rev}}^2/3)(C_{12}) \\
&\quad - v_{\mathrm{rev}}^2/6\} \text{ or}
\end{aligned}
$$

$$
\begin{aligned}
&3(v_{\mathrm{rev}}^2/6)^2 - 2(v_{\mathrm{rev}}^2/6)(A_1 + A_2 - C_{12}) + (A_1 \cdot A_2 - C_{12}^2) \\
&\quad = 0,
\end{aligned}
$$

from which v_{rev}^2 can be solved.

$$
\begin{aligned}
A_1 \cdot A_3 - (C_{13})^2 = \{&v_x^2/2 + v_z^2/2 + v_x \cdot v_z + v_{\mathrm{rev}}^2/3\} \cdot \{v_x^2/2 \\
&+ v_z^2/2 - v_x \cdot v_z + v_{\mathrm{rev}}^2/3\} -
\end{aligned}
$$

$$
\begin{aligned}
\{-v_x^2/2 + v_z^2/2\} = &v_x^4/4 + v_x^2 \cdot v_z^2/4 - v_x^3 \cdot v_z/2 + (v_{\mathrm{rev}}^2/3) \\
&\times(v_x^2/2 + v_z^2/2 + v_x \cdot v_z + v_x^2/2 + v_z^2/2 \\
&- v_x \cdot v_z + v_{\mathrm{rev}}^2/3) +
\end{aligned}
$$

$$
\begin{aligned}
&v_x^2 \cdot v_z^2/4 + v_z^4/4 - v_x \cdot v_z^3/2 + v_x^3 \cdot v_z/2 + v_x \cdot v_z^3/2 - v_x^2 \cdot v_z^2 \\
&\quad - v_x^4/4 - v_z^4/4 + v_x^2 \cdot v_z^2/2 =
\end{aligned}
$$

$$
(v_{\mathrm{rev}}^2/3)(A_1 + A_3 - v_{\mathrm{rev}}^2/3) \text{ or}
$$

$$
4(v_{\mathrm{rev}}^2/6)^2 - 2(v_{\mathrm{rev}}^2/6)(A_1 + A_3) + (A_1 \cdot A_3 - C_{13}^2) = 0,
$$

from which v_{rev} can be solved.

An alternative way is as follows. Write $A_1=A_1+2\cdot R$, where $R=v_{\mathrm{rev}}^2/6$; $A2=A_2+2\cdot R$, $A3=A_3+2\cdot R$, $C_{12}=\sqrt{(A_1\cdot A_2)}+R$ and $C_{13}=\sqrt{(A_1\cdot A_3)}$.

$$
\begin{aligned}
A_1 \cdot A_2 - C_{12}^2 &= (A_1 + 2R)(A_2 + 2R) - [\sqrt{(A_1 \cdot A_2)} + R]2 \\
&= 2R\{A_1 + A_2 - \sqrt{(A_1 \cdot A_2)}\} + 3R^2.
\end{aligned}
$$

$A_1+A_2=A_1+A_2-4R$ and $\sqrt{(A_1\cdot A_2)}=C_{12}-R$, thus

$$
\begin{aligned}
A_1 \cdot A_2 - C_{12}^2 &= 2R\{A_1 + A_2 - 4R - C_{12} + R\} + 3R^2 \quad \text{or} \\
&- 3R^2 + 2R(A_1 + A_2 - C_{12}) = A_1 \cdot A_2 - C_{12}^2
\end{aligned}
$$

$A_1\cdot A_3-(C_{13})^2=A_1\cdot A_3+2R(A_1+A_3)+4R2-A_1\cdot A_3=2R(A_1+A_3-4R)+4R^2=-4R2+2R(A_1+A_3)$, which are the same as the above equations.

[1]M. J. Berliner and J. F. Lindberg, ed., "Acoustic particle velocity sensors: Design, performance, and applications," *AIP Conference Proceedings 368* (Woodbury, New York, 1995), ISBN 1-56396-549-6; K. Kim, T. Gabrielson, and G. C. Lauchle, "Development of an accelerometer-based underwater acoustic intensity sensor," J. Acoust. Soc. Am. **116**, 3384–3392

(2004).
[2]A. Nordby and O.-H. Bjor, *Proceedings of Internoise* (1984), p. 1107–1109; O.-H. Bjor and H. J. Krystad, *Proc. of the Autumn Conference 1982* (Anon. Institute of Acoustics, Edinburgh, 1982), pp. B7.1–B7.5.
[3]H-E de Bree, "An overview of Microflown Technologies," Acta. Acust. Acust. **89**, 163–172 (2003); The Microflown, ISBN 9036515793; www.microflown.com/
[4]R. Raangs, T. Schlicke, and Richard Barham, "Calibration of a micromachined particle velocity microphone in a standing wave tube using a LDA photon correlation technique," Meas. Sci. Technol.. **16**, 1099–1108 (2005).
[5]V. B. Svetovoy and I. A. Winter, "Model of the μ-flown microphone," Sens. Actuators **86**, 171–181 (2000).
[6]J. W. van Honschoten, W. F. Druyvesteyn, H. Kuipers, R. Raangs, and G. J. M. Krijnen, "Selfnoise reduction in acoustic measurements with a particle velocity sensor by means of a cross correlation technique," Acta. Acust. **90**, 349–355 (2004).
[7]H-E de Bree, R. Raangs, and W. F. Druyvesteyn, "Sound intensity measurements with the microflown sensor," *Proc. Internoise 2004*, Prague, session AN3, Paper 3, Fig. 3 (Microflown Technologies B.V., University of Twente, 2004).
[8]V. A. Gordienko, B. I. Goncharenko, and Ya. A. Ilyushin, "Basic rules of vector-phase structure formulation of the ocean noise field," Acoust. Phys. **39**, 237–242 (1993).
[9]V. A. Shchurov, *Modern State and Prospects for Use of Underwater Acoustic Intensity Measurements* (Pacific Oceanological Institute, Vladivostok, 1998).
[10]F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995), lSBN 0 419 19810 5.
[11]R. Raangs, W. F. Druyvesteyn, and H. E. de Bree, J. Audio Eng. Soc. **51**, 344–357 (2003); J. W. van Honschoten, G. J. M. Krijnen, V. B. Svetovoy, H. E. de Bree, and M. C. Elwenspoek, J. Micromech. Microeng. **14**, 1468–1477 (2004).

A cross-spectrum weighting algorithm for speech enhancement and array processing: Combining phase-shift information and stationary signal properties

Ingo Schwetz[a)]
Friedhofstrasse 13, D-74223 Flein, Germany

Gerhard Gruhler
Faculty of Electronics and Mechatronics, University of Applied Sciences of Heilbronn, Max-Planck-Strasse 39, D-74081 Heilbronn, Germany

Klaus Obermayer
Fakultät für Elektrotechnik und Informatik, Technische Universität Berlin, Franklinstrasse 28/29, D-10587 Berlin, Germany

(Received 12 January 2004; revised 31 August 2005; accepted 5 October 2005)

In this paper, a gain function for noise cancellation with a two-channel microphone array is presented. This gain function combines ideas from one- and multichannel algorithms. It is developed using a minimum mean square error estimator for the amplitude of the speech signal from the cross spectrum between two microphone signals. To consider speech pauses and the absence of spectral components of the speech, an extension of this gain function is presented. The performance of the overall gain function is shown in terms of the cancellation of (diffuse) driving noise as well as the cancellation of an interfering speech signal, both recorded in a car. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149767]

PACS number(s): 43.60.Fg, 43.60.Vx, 43.60.Dh, 43.60.Uv [DOS] Pages: 952–964

I. INTRODUCTION

Noise cancellation in the car environment for telecommunication purposes is one of the most challenging problems in signal processing, because the output signals are received and rated by a real human listener: Noise reduction is desired, but weak nonlinear distortions as well as low production costs (e.g., small number of microphones) is a criterion which is hard to meet. Several single-channel methods were introduced in the past. The spectral subtraction method (Boll, 1979), for example, is widely used and well known, but one key problem of this method is the residual artificial noise, also known as "musical tones." The processed speech sounds very unnatural, which is a result of a rapidly changing gain function and of wrong noise estimates. In order to avoid the musical tones, different strategies have been proposed. In Gustafsson *et al.* (2001), the variability of the gain function was reduced. In Gustafsson (1999) a psychoacoustically motivated method to mask the distortions of the noise using the speech signal was presented. Another basic approach to estimate a gain function was given in Ephraim and Malah (1984), where the amplitude of the short time power spectra are found by a minimum mean square error (MMSE) estimator. Estimating if the additive noise is destructive or constructive and applying different weights for these cases is the basic idea in Soon and Koh (2000).

All the above-mentioned single-channel methods are dependent on an exact estimation of the noise spectrum. In Martin (2001) and Cohen (2003) this was done, for example, by using a recursive averaging of the noise spectrum based on minima statistics. As long as the noise is stationary over time, musical tones are well suppressed by state-of-the-art single-channel methods. As soon as the noise is nonstationary, however, nonlinear distortions arise. An approach for noise cancellation, which avoids the problem of musical tones, is based on beamforming with microphone arrays. The method of beamforming with antenna arrays was introduced more than half a century ago (Gilbert and Morgan, 1955), and beamforming with microphone arrays is receiving more and more attention (Brandstein and Ward, 2001). The basic filter-and-sum system consists of a linear finite impulse response filter for each microphone channel and a summation afterwards. Many methods for the estimation of these filters were presented in the past (e.g., Dörbecker, 1997; Seltzer and Raj, 2003; Dahl and Claesson, 1999), and also the combination with a Wiener postfiltering was considered (Marro *et al.*, 1998). Nonlinear filtering methods were also suggested (Knecht *et al.*, 1995). A different approach is used by the Generalized Sidelobe Canceller or Griffith-Jim Beamformer (Griffiths and Jim, 1982), where the speech signal is enhanced by subtracting filtered noise-only signals from the beamformer output, which are outputs of the so-called "blocking matrix." The beamforming approach is of key interest, not only for the area of speech enhancement in hands-free-kits, but also for hearing aids (Vanden Berghe and Wouters, 1998; Greenberg *et al.*, 2003) or conferencing systems (Meyer, 2001). As beam-forming is a slowly adapting linear filtering technique, interfering sources with very fast chang-

[a)]Free Researcher, working toward Ph.D. with the Technische Universität Berlin during the time of this work; electronic mail: ingo.schwetz@gmx.de

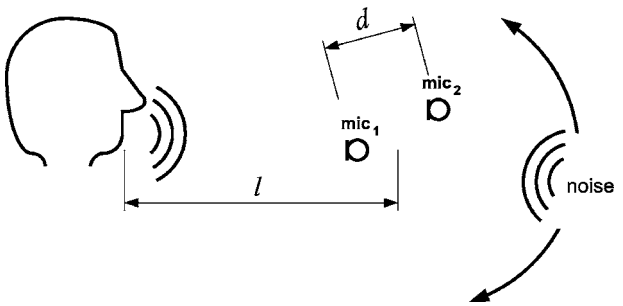


FIG. 1. Microphone array with two microphones.

ing positions could not be suppressed perfectly. This is also true for diffuse noise fields for which the "virtual" direction of arrival changes very fast.

To overcome the problem of rapidly changing directions of arrival, we suggest using a two-channel noise reduction method. It is based on the following ideas: Like in single-channel methods it is assumed that the amplitude of the short-time spectrum contains most of the information about the speech signal. Combining an estimate of the direction of incidence (in terms of the phase of the short-term cross spectrum), which is also the basic idea of the gain function presented in Aarabi and Shi (2004), with the long-term signal properties (estimates of the noise- and speech-power-average), a MMSE estimator to estimate the amplitude of the speech portion is derived. Then a gain function is constructed, which is used as a weight for the spectral amplitudes. As one will see, there is a strong relationship to the approach of Ephraim and Malah (1984), who introduced a MMSE estimator for the speech content in a short-term spectrum. It is also somewhat similar to the approach described in Lotter *et al.* (2003).

II. PROPERTIES OF A SHORT-TIME CROSS SPECTRUM BETWEEN TWO MICROPHONE SIGNALS

We consider a microphone array which consists of two microphones at arbitrary positions in the far field of a speaker (Fig. 1). We assume that the distance l between the speaker and the microphones is much larger than the size d of the array, so that the amplitude of the speech signal is approximately equal at both microphones. The speech signals from the speaker are disturbed by a diffuse noise field with the same amplitude properties, but with changing direction of arrival. Short time stationarity for the speech signal as well as for the noise signal is further assumed.

The short-time cross spectrum between the microphone signals in the frequency domain, $X_1(f)$ and $X_2(f)$, is given by

$$\mathbf{E}\{X_1^*(f)X_2(f)\} = Y = Xe^{j\psi} = Se^{j\varphi} + Re^{j\zeta}, \tag{1}$$

where the frequency dependency was dropped for convenience on the right-hand side. $\mathbf{E}\{\cdot\}$ denotes an expectation, and $Xe^{j\psi}$ the current short-time cross spectrum with squared amplitude X and phase ψ. The short-time cross spectrum is divided into a signal part $Se^{j\varphi}$ and a noise part $Re^{j\zeta}$, where S and R are the squared amplitude and φ and ζ the phases (see Fig. 2). Because the speaker remains at a fixed location, the phase φ is assumed to remain constant and depends on the

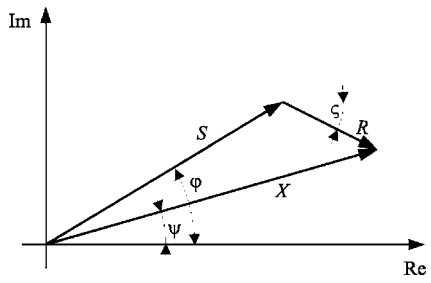


FIG. 2. Separation of the cross spectrum into a signal and a noise term.

delay of the speech signal between the two microphones as well as on phase shifts, e.g., due to neighboring reflecting surfaces. Figure 3, which shows the short-time cross spectrum of the recording of a real-world speech-only signal at 2 kHz and a microphone distance of 2.15 cm for different time points, verifies this assumption. The phase of the noise part ζ, however, fluctuates, because for a diffuse noise field all directions of arrival are equally probable. This is shown in Fig. 4 for a real-world recording of a noise-only signal at the same frequency and using the same microphone setup as for the speech signal. Using these properties, we now develop an estimator for the amplitude S of the speech-signal given the amplitude X and the phase ψ of the current cross spectrum.

III. MINIMUM MEAN SQUARE ERROR ESTIMATOR

In order to estimate the squared amplitude S of the speech signal from the squared amplitude X and the phase ψ of the current cross spectrum, we apply Bayes' theorem for densities. We obtain (Papoulis, 1991):

$$\hat{S} = \mathbf{E}\{S|X,\psi\} = \mathbf{E}\{S|Y\} = \frac{\int_0^\infty \int_0^{2\pi} Sp(Y|S,\varphi)p(S,\varphi)d\varphi dS}{\int_0^\infty \int_0^{2\pi} p(Y|S,\varphi)p(S,\varphi)d\varphi dS}, \tag{2}$$

where $\hat{S}$ is the estimated squared amplitude of the speech

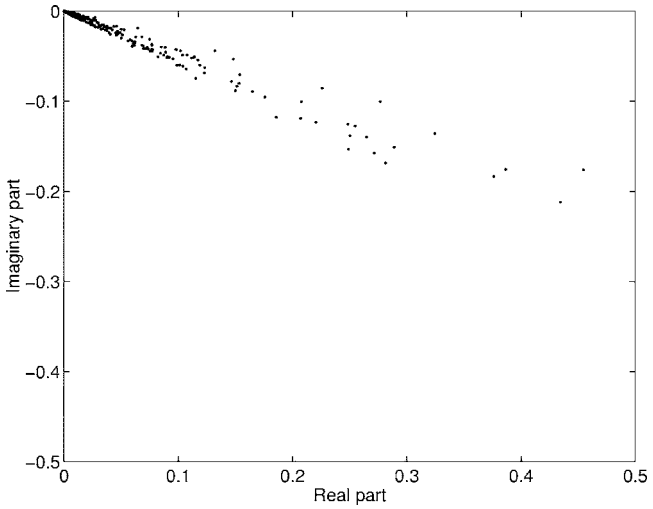


FIG. 3. Short time cross spectrum at different points in time for a real-world speech-only recording (microphone distance d=2.15 cm, 512-point-FFT, 8 kHz sampling frequency, result for frequency bin 128, f=2000 Hz).

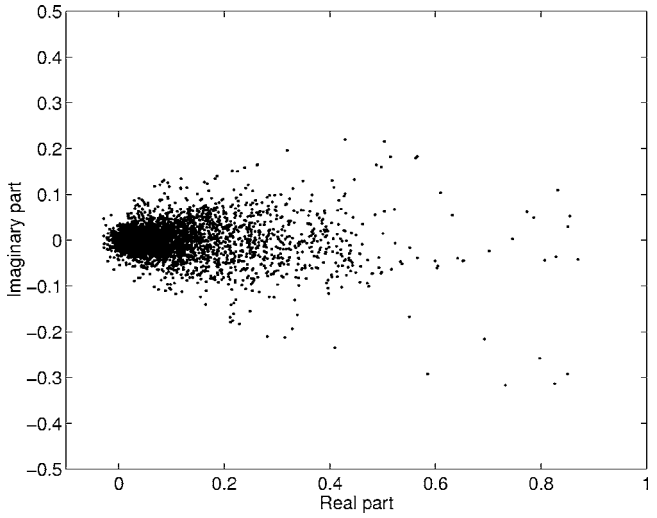


FIG. 4. Short time cross spectrum at different points in time for a real-world noise-only recording (microphone distance d=2.15 cm, 512-point-FFT, 8 kHz sampling frequency, result for frequency bin 128, f=2000 Hz).

signal and $p(\cdot)$ is a probability density function (pdf). In order to evaluate this equation, a model for the pdf $p(Y|S,\varphi)$ of the current cross-spectrum Y has to be found under the assumption that the speech amplitude S with phase φ is present. Additionally, the pdf $p(S,\varphi)$ of the speech signal is needed.

IV. STATISTICAL MODEL

We now evaluate Eq. (2) and define proper statistical models for the speech and the noise signals. Because the noise signal is assumed to be diffuse, the amplitude R and the phase ζ of the noise are assumed to be independent of each other. The pdf of the cross-spectrum Y conditioned on the speech signal $Se^{j\varphi}$ is then given by

$$p(Y|S,\varphi) = p(X,\psi|S,\varphi) = \underbrace{p(|Xe^{j\psi} - Se^{j\varphi}|)}_{p(R)}\,\underbrace{p(\arg(Xe^{j\psi} - Se^{j\varphi}))}_{p(\zeta)}. \quad (3)$$

If the short-time amplitude $r=\sqrt{R}$ of the noise is assumed to be Gaussian, then the pdf of R is given by (Papoulis, 1991):

$$p(R) = \frac{1}{\sqrt{2\pi R\lambda_r}} \exp\left(-\frac{1}{2\lambda_r}R\right), \quad (4)$$

with the variance of r given by λ_r. Under the assumptions of a diffuse noise field and short time stationarity, the pdf of the phase ζ is given by the uniform density (see Appendix A)

$$p(\zeta) = \frac{1}{4\pi}\frac{c}{fd}, \quad \zeta = -2\pi\frac{fd}{c} \ldots 2\pi\frac{fd}{c}, \quad (5)$$

where c is the speed of sound, f denotes the frequency, and d is the microphone distance. The distribution given in Eqs. (3)–(5) is shown in Fig. 5, where the gray area represents the pdf, dark and bright areas indicate high and low values. The sharp edges are due to the minimum and maximum phase given in Eq. (5).

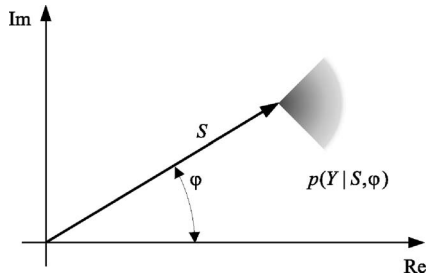


FIG. 5. Probability density function of the noise.

Unfortunately, Eq. (2) with Eq. (4) and Eq. (5) cannot be solved in closed form. [1] Therefore, we approximate the pdfs as follows. For the real and the imaginary parts of Y, assuming that speech is present, we chose the jointly normal distribution given by (Papoulis, 1991), Eqs. 6–15)

$$\begin{aligned} p(Y|S,\varphi) &= p(\mathrm{Re}(Y),\mathrm{Im}(Y)|S,\varphi) \\ &= \frac{1}{\sqrt{2\pi\lambda_{\mathrm{Re}}}}\exp\left(-\frac{1}{2\lambda_{\mathrm{Re}}}|X\cos(\psi) - S\cos(\varphi_0) - \kappa|^2\right) \\ &\quad \times \frac{1}{\sqrt{2\pi\lambda_{\mathrm{Im}}}}\exp\left(-\frac{1}{2\lambda_{\mathrm{Im}}}|X\sin(\psi) - S\sin(\varphi_0)|^2\right), \end{aligned} \quad (6)$$

where λ_{Re} and λ_{Im} are the variances of the real and imaginary part of the noise, $X\cos(\psi)-S\cos(\varphi_0)$ is the real part of the noise, $X\sin(\psi)-S\sin(\varphi_0)$ is the imaginary part of the noise and where we additionally assumed independence of the real and imaginary parts—in order to obtain an analytical solution. In addition, for an eventually better match of the pdfs in Eqs. (4) and (5), the pdf can be shifted along the real axis by using the variable κ. Appendix B provides a method for choosing a good value for this variable. The symmetry of the imaginary part of Eq. (3) given by Eq. (5) shows that such a parameter is not necessary for the imaginary part of Eq. (6). The phase φ of the signal's cross spectrum is always constant and is denoted by φ_0. This distribution is shown in Fig. 6, where the mean value is marked by a cross, and the variances of the real and imaginary parts are depicted by the dashed line. The maximum phase angle of the diffuse noise $\zeta_{\max}$, which is dependent on the microphone distance, is also shown.

In order to achieve a closed form solution, the pdf of the squared amplitude of the speech signal is not modeled as a Laplacian distribution, but as a Gaussian distribution:

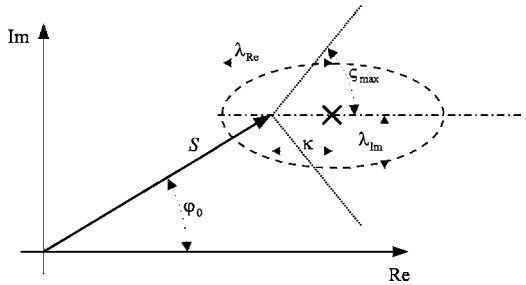


FIG. 6. Model of the pdf of the noise.

$$p(S,\varphi) = \frac{1}{\sqrt{2\pi\lambda_S}} \exp\left(-\frac{1}{2\lambda_S} S^2\right). \tag{7}$$

V. MMSE GAIN FUNCTION AND MODELING SPEECH UNCERTAINTY

Using Eqs. (6) and (7), an analytical solution of Eq. (2) can be obtained. In Appendix C we show that

$$\hat{S} = \underbrace{\frac{\sqrt{\upsilon}}{\gamma}\frac{D_{-2}(\Theta)}{D_{-1}(\Theta)}}_{=:G} X, \tag{8}$$

where υ, ξ, γ, and Θ are given by Eqs. (C6)–(C9). The functions $D_p(z)$ are the parabolic cylinder functions, Eq. (C4), hence the final MMSE gain function can be written as

$$\begin{aligned} G &= \frac{\sqrt{\upsilon}}{\gamma}\frac{1}{\sqrt{2}}\frac{2\Phi\left(1,\frac{1}{2};\frac{\Theta^2}{2}\right) - \sqrt{2\pi}\Theta\Phi\left(\frac{3}{2},\frac{3}{2};\frac{\Theta^2}{2}\right)}{\sqrt{\pi}\Phi\left(\frac{1}{2},\frac{1}{2};\frac{\Theta^2}{2}\right) - \sqrt{2}\Theta\Phi\left(1,\frac{3}{2};\frac{\Theta^2}{2}\right)} \\ &= \frac{\sqrt{\upsilon}}{\gamma}\frac{1}{\sqrt{2}}\frac{2\Phi\left(1,\frac{1}{2},\frac{\Theta^2}{2}\right) - \sqrt{2\pi}\Theta\exp\left(\frac{\Theta^2}{2}\right)}{\sqrt{\pi}\exp\left(\frac{\Theta^2}{2}\right) - \sqrt{2}\Theta\Phi\left(1,\frac{3}{2},\frac{\Theta^2}{2}\right)}, \end{aligned} \tag{9}$$

where $\Phi(\alpha,\beta;z)$ denotes the hypergeometric function given by Eq. (C5).

The gain function shown in Eq. (9) was derived under the assumption that speech is present. In real speech signals, this assumption does not hold, because in a natural speech segment, there are many speech pauses as well as frequency bins without a speech content, e.g., due to the harmonic structure of voiced speech, which results in a comblike spectrum. If speech is assumed to be present at any time, the gain function amplifies low amplitudes in the speech pauses. This can be verified using Eq. (9), where $X \to 0$ leads to $\sqrt{\upsilon}/\gamma \to \infty$. Therefore, the gain function Eq. (9) cannot be used without further modifications, because it would cause noise artifacts in sections where signal power is low. In Ephraim and Malah (1984) the following modification was derived.

$$\hat{G} = \frac{\Lambda}{1+\Lambda} G, \tag{10}$$

where Λ is the generalized likelihood ratio

$$\Lambda = \mu \frac{p(Y|H_1)}{p(Y|H_0)} \tag{11}$$

of the probabilities for the two hypothesis H_1, for speech being present, and H_0, for a speech pause. The factor μ is defined as $\mu=(1-q)/q$, where q is the probability of the speech pause. With the two pdfs for the two hypotheses

$$\begin{aligned} p_d(Y|H_1) = &\frac{1}{\sqrt{2\pi\lambda_{\mathrm{Re}}}} \\ &\times \exp\left(-\frac{1}{2\lambda_{\mathrm{Re}}}|X\cos(\psi) - \sqrt{\lambda_S}\cos(\varphi_0) - \kappa|^2\right) \\ &\times \frac{1}{\sqrt{2\pi\lambda_{\mathrm{Im}}}} \exp\left(-\frac{1}{2\lambda_{\mathrm{Im}}}|X\sin(\psi) - \sqrt{\lambda_S}\sin(\varphi_0)|^2\right) \end{aligned} \tag{12}$$

for the active speech, where the amplitude of the speech is assumed as $\sqrt{\lambda_S}$, and

$$\begin{aligned} p_d(Y|H_0) = &\frac{1}{\sqrt{2\pi\lambda_{\mathrm{Re}}}} \exp\left(-\frac{1}{2\lambda_{\mathrm{Re}}}|X\cos(\psi) - \kappa|^2\right) \\ &\times \frac{1}{\sqrt{2\pi\lambda_{\mathrm{Im}}}} \exp\left(-\frac{1}{2\lambda_{\mathrm{Im}}}|X\sin(\psi)|^2\right) \end{aligned} \tag{13}$$

for the speech pause, the generalized likelihood ratio Eq. (11) is given by

$$\begin{aligned} \Lambda = \mu \exp\Bigg(&\frac{1}{\lambda_{\mathrm{Re}}}\left(X\sqrt{\lambda_S}\cos(\psi)\cos(\varphi_0) - \frac{\lambda_S}{2}\cos^2(\varphi_0)\right. \\ &\left. - \sqrt{\lambda_S}\cos(\varphi_0)\kappa\right) + \frac{1}{\lambda_{\mathrm{Im}}}\left(X\sqrt{\lambda_S}\sin(\psi)\sin(\varphi_0)\right. \\ &\left.\left. - \frac{\lambda_S}{2}\sin^2(\varphi_0)\right)\right). \end{aligned} \tag{14}$$

Figures 7–9 show the gain-function $\hat{G}$ for different parameters. As the gain function approaches very low values for certain angles, e.g., $\psi<0$, the plot is scaled to $-20,\ldots,0$ dB. The angle φ_0 of the main incidence as well as the probability q of the speech pause are always fixed to the same value. From Figs. 7–9 the signal-to-noise ratio (SNR) is increasing. If the SNR is small (Fig. 7), the gain is small for high values of the amplitude X and high for the low values. In the case of high SNRs as in Fig. 9, the gain function is only high for high amplitudes X, so low amplitudes are treated as noise only. As one can see in any of the figures, the phase ψ is always playing a key role for the gain. Values of ψ which are obviously a result of high noise cause low gain values. Note that the highest gain values are not located at $\psi=\varphi_0$, which is explained in Appendix D.

A signal incident from an interfering source, located at a symmetrical position to the speech source (as in Fig. 10), causes a phase angle ψ with a different sign than φ_0 and is, therefore, suppressed by the low gain values. Reflections of the speech signal itself, or an interfering source signal causing phase angles $\psi<\varphi_0$, will also be reduced. Therefore, the gain function will cause a dereverberation as well as a suppression of interfering sources.

To avoid distortions caused by aggressively scaled gain values, we introduce the gain limitation

$$\hat{G}' = \alpha + (1-\alpha)\hat{G}, \quad 0 \leqslant \alpha \leqslant 1. \tag{15}$$

For $\alpha=0$ the gain function remains unchanged and maximum noise reduction is performed; for $\alpha=1$ the gain function

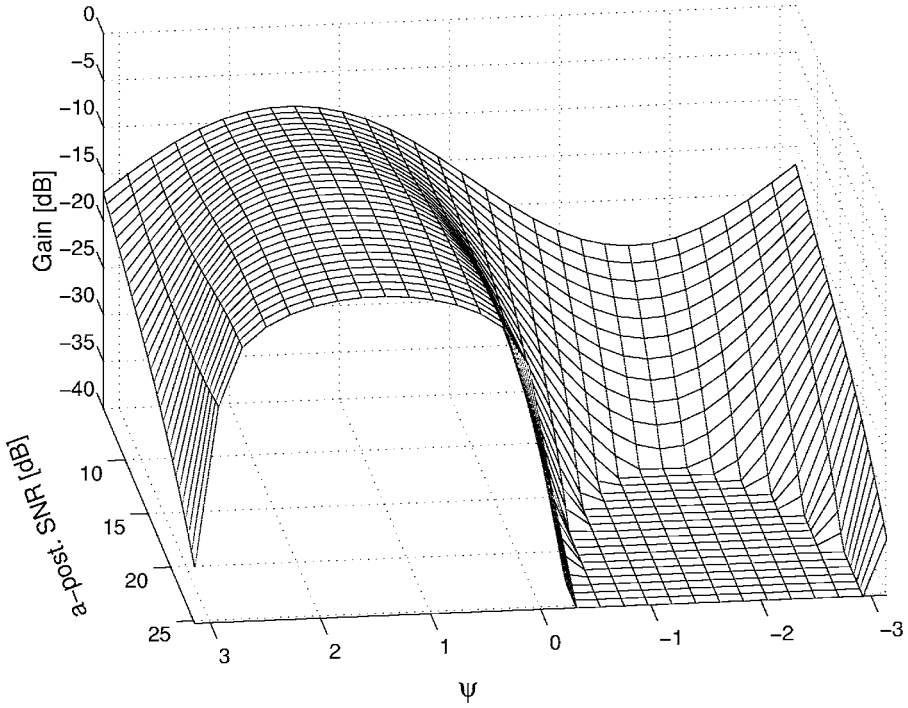


FIG. 7. Gain as a function of the signal phase Ψ and the signal power X (described by the *a posteriori* SNR) for $\sqrt{\lambda_S}=0.02$, $\sqrt{\lambda_{\mathrm{Re}}}=0.1$, and $\sqrt{\lambda_{\mathrm{Im}}}=0.03$ (*a priori* SNR=−8.7 dB), $\varphi_0=\pi/4$, $\kappa=0.07$, and $q=0.8$.

equals unity and the signals remain unchanged. Table I describes what $\hat{G}'$ depends on and how these values can be estimated.

VI. TEST SETUP

In order to assess the performance of the proposed method in a real world scenario we used the following tests. A male and a female speaker were seated at the driver's and the passenger's seat of a Volkswagen car (Golf III station wagon, reverberation time $T_{60}=50$ ms). Two omnidirectional microphones were mounted with a spacing of 2.15 cm above the interior mirror at the roof of the car. This setup with the corresponding directions of incidence is depicted in Fig. 10. The used spacing represents a distance, which is equal to about 1/4 of the wavelength at 4 kHz. Therefore $|\zeta_{\max}| \approx \pi/2$. Larger spacings would result in a situation similar to the "spatial aliasing" (Brandstein and Ward, 2001) of beamforming algorithm for high frequencies, where one specific phase shift between the microphone signals is no longer unique for one direction of arrival and sidelobes occur.

To obtain clean recordings of the speech and noise parts, each speaker as well as the driving noise was recorded separately. Prior to the performance analysis, the phase angles $\varphi_0|_{\mathrm{male}}$ for the male speaker and $\varphi_0|_{\mathrm{female}}$ for the female

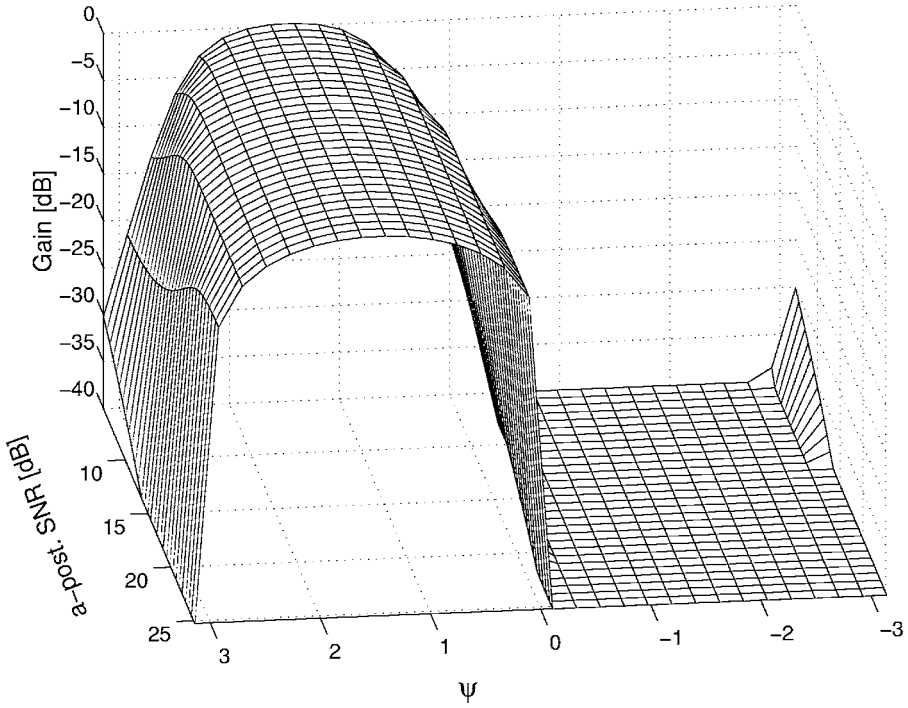


FIG. 8. Gain as a function of the signal phase Ψ and the signal power X (described by the *a posteriori* SNR) for $\sqrt{\lambda_S}=0.1$, $\sqrt{\lambda_{\mathrm{Re}}}=0.1$, and $\sqrt{\lambda_{\mathrm{Im}}}=0.03$ (*a priori* SNR=5.2 dB), $\varphi_0=\pi/4$, $\kappa=0.07$, and $q=0.8$.

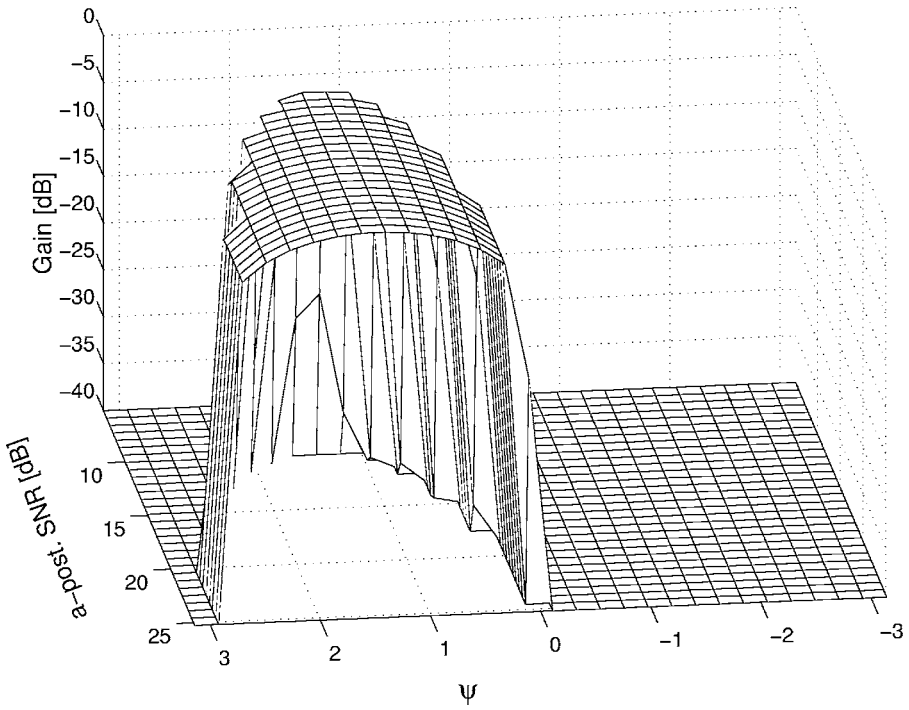


FIG. 9. Gain as a function of the signal phase Ψ and the signal power X (described by the *a posteriori* SNR) for $\sqrt{\lambda_S}=0.5$, $\sqrt{\lambda_{\mathrm{Re}}}=0.1$, and $\sqrt{\lambda_{\mathrm{Im}}}=0.03$ (*a priori* SNR=19.2 dB), $\varphi_0=\pi/4$, $\kappa=0.07$, and $q=0.8$.

speaker were measured and used as the phase information input for the presented simulations. The results of these measurements are shown in Fig. 11. The figure shows an almost linear relationship between the phase and the frequency, which is a result of the transport time of the signal from the one microphone to the other. [2]

The components of the algorithm used to filter the signals with the proposed gain function are shown in Fig. 12. The two discrete-time microphone signals $x_1(k)$ and $x_2(k)$ are analyzed with a windowing function (Hanning) and are transformed to the frequency domain using the fast Fourier transform (FFT). Afterwards the cross spectrum is calculated by conjugation of one signal and subsequent multiplication. The signal processing path is as follows: The absolute value of the amplitude of the cross spectrum is estimated and multiplied by the gain $\hat{G}$ in each frequency band. To get the output amplitude of each spectral component of the signal, the square root is taken. The phase of the output signal is taken from the phase of one of the input signals [in our case from signal $X_1(f)$]. The output signal $y(k)$ is then reconstructed using the inverse FFT and the overlap-and-add method (Rabiner and Schafer, 1978). In order to achieve an exact evaluation of the performance of the algorithm, the speech signals and noise signals were separately filtered using this approach. Therefore, a separate output for each of the input signals was available.

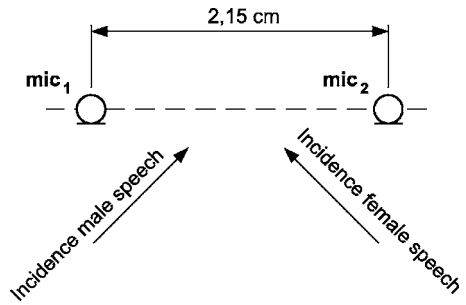


FIG. 10. Microphone array with directions of incidence of the two speech signals.

TABLE I. Parameters of $\hat{G}'$ and their estimation.

Variable	Description	Estimation
X	Amplitude of short-time cross spectrum	Estimation of short time cross spectrum, e.g., by recursive averaging
ψ	Phase of cross spectrum in current frame	Estimation of short time cross spectrum, e.g., by recursive averaging
φ_0	Assumed phase of cross spectrum for speech signal	Measuring phase of cross spectrum using Eq. (1) when only speech is present; result as in Fig. 11
λ_{Re}	Variance of the real part of noise-only cross spectrum	From short-time cross spectrum and, e.g., using Eq. (16)
λ_{Im}	Variance of the imaginary part of noise-only cross spectrum	From short-time cross spectrum and, e.g., using Eq. (16)
λ_S	Variance of the amplitude of the speech-only cross spectrum	From short-time cross spectrum and, e.g., using Eq. (17)
λ	Evaluated from φ_0, λ_{Re}, λ_{Im}, and λ_S.	Eq. (C2)
Θ	Evaluated from φ_0, X, λ, λ_{Re}, λ_{Im}	Using Eq. (C9)
v, λ	*A priori* SNR and *a posteriori* SNR	Using Eqs. (C6)–(C8)
κ	Shift parameter for matching the noise	Either set to zero or by estimation from noise variance following Appendix B
q	Probability of a speech-pause	Either by an ad hoc chosen fixed value or (better) by using a speech/pause detector
α	Gain-normalization factor	Ad hoc chosen fixed value

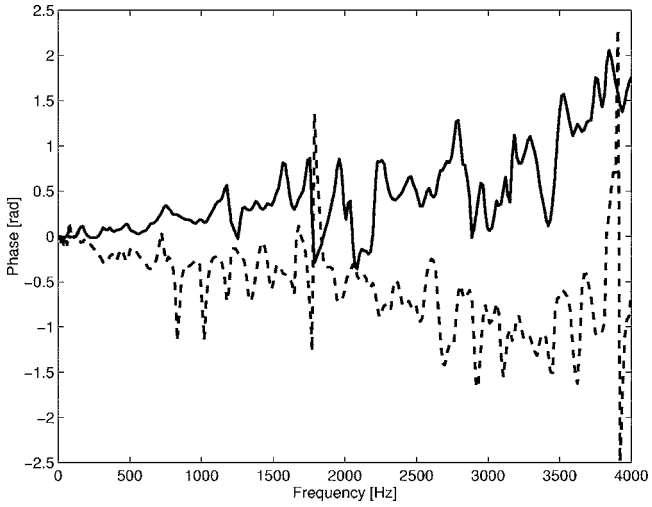


FIG. 11. Measured phase $\varphi_0|_{\text{male}}$ for the male speaker (solid line) and the female speaker $\varphi_0|_{\text{female}}$ (dotted line).

To calculate the gain function for each frequency component, the amplitude X and the phase ψ of the cross spectrum, the phase φ_0 regarding the expected direction of incidence of the desired speech, as well as the *a priori* variances of the real (λ_{Re}) and the imaginary (λ_{Im}) parts of the noise and the speech (λ_S) signals must be estimated. If an independent measurement of the noise components is not possible, approaches like minima-tracking (Martin, 2001) can also be used. If the microphones are placed in an environment with free field (or close to free field) conditions, φ_0 can be estimated by geometric considerations [using Eq. (A4) and replacing ζ by φ_0, where θ represents the direction of incidence]. Another possibility is to measure the phase of the cross spectrum, Eq. (1), when only the desired signal is present; this approach was used for the presented results. For the estimation of λ_{Re} and λ_{Im} at time-frame ν and for frequency bin q we used (Doblinger, 1995):

$$\lambda_{\text{Re}}(q,\nu) = \alpha_{\text{Re}}\lambda_{\text{Re}}(q,\nu-1) + (1-\alpha_{\text{Re}})(\text{Re}(Y(q,\nu)) - \text{Re}(Y(q,\nu-1)))^2,$$

$$\lambda_{\text{Im}}(q,\nu) = \alpha_{\text{Im}}\lambda_{\text{Im}}(q,\nu-1) + (1-\alpha_{\text{Im}})(\text{Im}(Y(q,\nu)) - \text{Im}(Y(q,\nu-1)))^2, \quad (16)$$

where $Y(q,\nu)$ is the cross spectrum [Eq. (1)] at time-frame ν and frequency bin q. The estimate of the *a priori* variance of

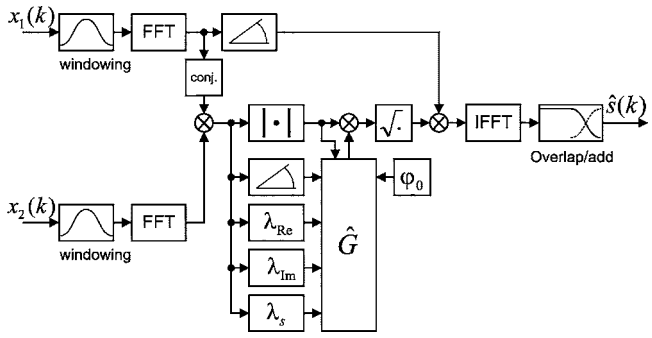


FIG. 12. Simulation setup of the algorithm for filtering with the proposed gain function.

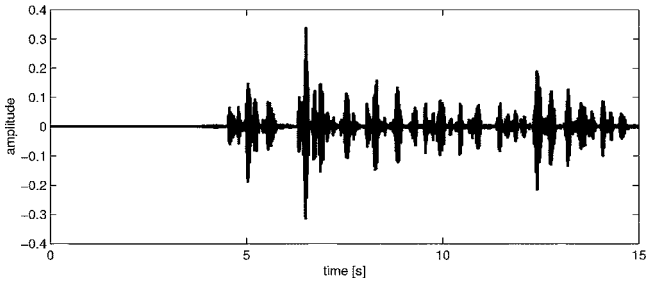


FIG. 13. Input signal of the female speaker from the passenger's seat (left channel only).

the speech signal λ_S was obtained by a simple recursive average of the amplitude of the cross spectrum,

$$\lambda_S(q,\nu) = \alpha_S\lambda_S(q,\nu-1) + (1-\alpha_S)|Y(q,\nu)|^2. \quad (17)$$

The values from Eqs. (16) and (17) were then used to calculate the gain function.

Throughout the tests, the length of the windowing function of the algorithm was 512 samples, shifted by 256 samples at each step (overlap was 50%). For the value κ in Eq. (C9) the result from Appendix B was used, so $\kappa = \sqrt{\frac{1}{2}\lambda_{\text{Re}}}$. The parameters for the noise averaging were chosen ad hoc to be: $\alpha_{\text{Re}}=\alpha_{\text{Im}}=0.95$, $\alpha_S=0.1$ and the probability for the absence of speech was set to $q=0.95$, which gave a good overall result in terms of the noise reduction evaluated by listening tests.

The following test protocol was used: Speech sequences of the male speaker from the driver's seat and the female speaker from the passenger's seat as well as driving noise were recorded separately. As the filtering of the signals was done separately, the calculation of the input and output SNR and signal-to-interference ratio (SIR) was possible. The SNR numbers presented here were calculated by averaging the signal's power only in segments with an active speech signal. The signals in speech pauses were not considered.

The recordings were done with 16 bit precision and at a sampling frequency of 8 kHz; the signals were high-pass filtered with a cutoff frequency of 300 Hz in order to achieve signals within the telephone frequency band. The input files to the algorithm were handled in the following way:

(1) The female speech signal was scaled to achieve 0 dB SNR in the right channel of the raw input signal.
(2) The male speech signal was scaled to achieve 0 dB SNR in the left channel of the raw input signal.
(3) All three signals were added, while scaling the noise to achieve different SNRs.

Figure 13 depicts the female speech signal from the left microphone, Fig. 14 the male speech signal as well as Fig. 15 the noise signal from the same microphone. As one can see, there is a segment in the middle of the signals, where the female and male speaker signals overlap.

VII. SIMULATION RESULTS

The performance estimation was done by using a measure for the enhancement of the signal-to-noise ratio (to assess the noise-cancellation capabilities) and by using a mea-

sure for the enhancement of the signal-to-interference ratio (to assess the cancellation properties regarding an interfering source). The result in terms of noise cancellation was calculated as the signal-to-noise-ratio-enhancement (SNRE) and is given as the difference between the SNR after noise cancellation and the SNR of the input signal. The power of the speech and noise signals to calculate the values of the SNRs were estimated only over the segments containing speech. In order to achieve an exact performance evaluation, the two different speech signals and the noise signals were filtered separately.

For comparison purposes, the same signals were filtered using different algorithms: "Proposed$_{-\infty}$" and "Proposed$_{-30}$" give the results for the proposed gain function with the gain-limiting parameter α in Eq. (15) set to $\alpha=0$ (no gain limit) and $\alpha=0.0316$ (gain limit of -30 dB). "Aarabi$_{\gamma=50}$" gives the result from using the gain function based on a phase weighting described in Aarabi and Shi (2004) with the constant γ, which is setting the "aggressiveness" of the gain function, set to a value of 50. This value was chosen from listening tests to achieve a good performance in terms of the SNRE/SIRE while musical tones were in the order of the "Proposed$_{-30}$" method. "SDBF" and "DSBF" gives the results from superdirective-beamforming and a delay-and-sum-beamforming.

Figures 16 and 17 show the performance in terms of the SNRE for the different algorithms for different input SNRs. The proposed method (with and without gain limitation) performed best in terms of the achieved SNRE over the whole range of input SNRs. Because the performance of the DSBF was of the order of the SNR differences between the two channels (in most of the cases even slightly negative), the results were set to 0. This result can be justified by the very small distance between the two microphones that makes the DSBF method ineffective.

The result in terms of canceling an interfering talker was calculated in the following way: The undesired speech (female speech in the left channel, male speech in the right channel) is defined as interference, while the desired speech (male speech in the left channel, female speech in the right channel) is termed as signal. With this definition, the signal-to-interference-ratio-enhancement (SIRE) was calculated using the same procedure as mentioned earlier. Figures 18 and 19 are showing the results for different input SNRs, while the input SIRs were kept constant at 0.1 dB for the right channel and 0.7 dB for the left channel. The proposed method performed well, however, the achieved SIRE was

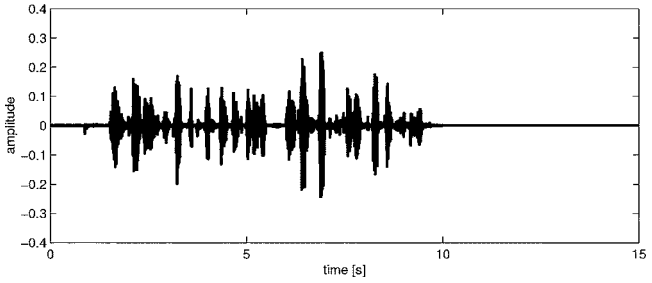


FIG. 14. Input signal of the male speaker from the driver's seat (left channel only).

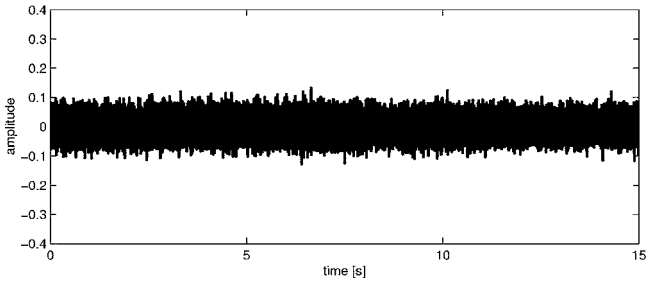


FIG. 15. Input driving noise signal (left channel only).

comparable to the SIRE achieved by the arabi method. One reason for the approximately equal performance is that the interfering speech signal is not modeled in the proposed approach and that it is highly dependent on the direction of incidence. Also, the gain function of the arabi method more aggressively suppresses frequency bins where the phase of the cross spectrum deviates from the "desired" phase. This is also the reason for the different performance for the male and the female signals, as even slight deviations in the phase results in highly different gain values. In high noise conditions (low SNR), the performance of the Aarabi algorithm rapidly drops. The performance of the DSBF was again very weak and was on the order of the SIR differences between the two channels (again, in most of the cases even negative), so that the results were again set to 0. As earlier, this result can be justified by the very small distance between the two microphones and the resulting ineffectiveness of the DSBF.

Listening tests show the following result: After a short transient, the signals filtered with "Proposed$_{-30}$" and "Aarabi$_{\gamma=50}$" contain some low amplitude musical tones of about the same strength, while DSBF and SDBF do not suffer from these distortions. The signals filtered by the Aarabi method, however, suffer from a loss of high frequency speech content and from reverberations, which make the overall signal quality less pleasing. This loss of speech quality was noticed for all different input SNRs. The signals fil-

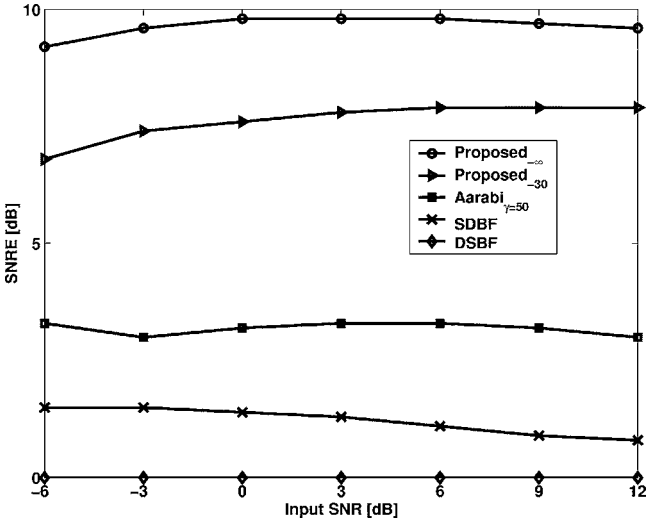


FIG. 16. Noise reduction performance for the driver (male speaker) for different input SNRs.

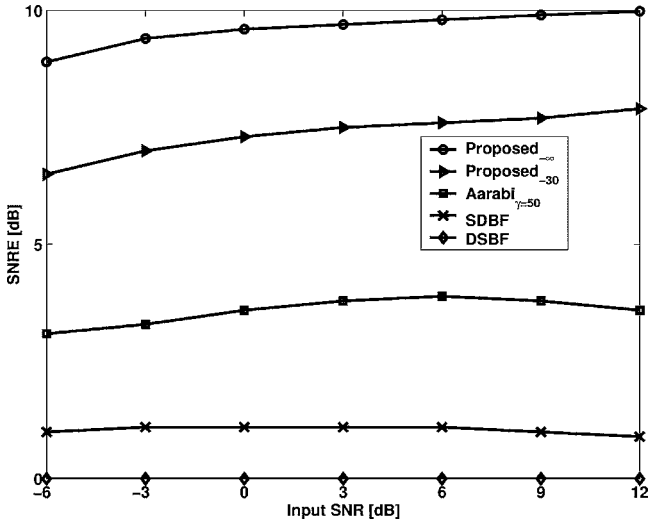


FIG. 17. Noise reduction performance for the passenger (female speaker) for different input SNRs.

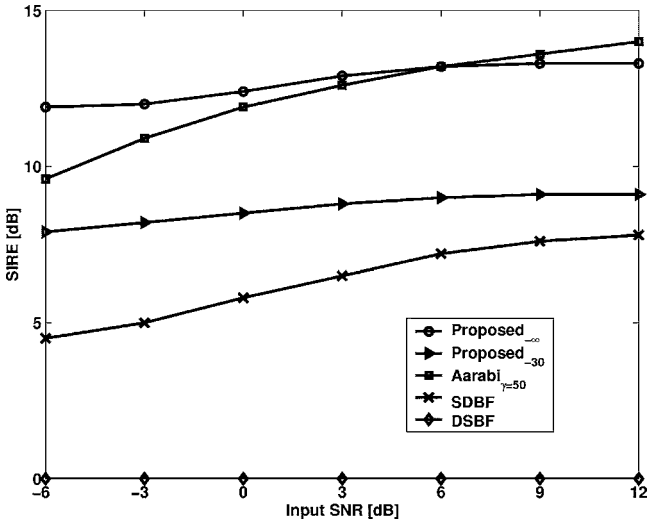


FIG. 19. Canceling of the interference for the passenger input (female speaker) for different input SNRs. The input SIRs were kept constant.

tered by the fixed beamforming algorithms sound natural, but suffer from a high noise content and a low interference suppression.

To estimate the susceptibility against errors in the estimated phase, the following test was performed with the algorithm "Proposed$_{-\infty}$:" The same signals as above were used with a SNR of 0 dB. As an input to the algorithm, a phase error, linearly increasing with the frequency, was added to the phase input described earlier, so the error at $f=0$ Hz was always 0 rad and the error at $f=4$ kHz was in the range of -0.4 to 0.4 rad. Figure 20 shows that the SNRE is not very sensitive against this kind of error for the used setup. The effect for negative errors (1.6 dB at -0.4 rad) is slightly larger than the effect for positive errors (0.6 dB at +0.4 rad).

The results show that the proposed method can simultaneously suppress diffuse noise and an interference from another localized sound source. Compared to two-channel superdirective-beamforming algorithms, the ability to suppress noise and to cancel interfering speech signals is superior but the amount of signal distortions is increased. In contrast to single-channel methods like spectral subtraction, the current method is capable not only of suppressing noise but also canceling unwanted interferences.

VIII. CONCLUSIONS AND FUTURE WORK

The proposed gain function is capable of noise cancellation as well as performing cancellation of an interfering signal by using a microphone array with two microphones. This was shown by simulation results done with real world recordings.

The presented results were based on very rough estimations of the average noise and speech power. With better estimations of these values, the performance of the algorithm can surely be improved in terms of noise/interference reduction as well as in terms of signal distortions and musical tones. Another possibility to enhance the performance would be to improve the estimation of speech pauses, e.g., by using a speech/pause detector. Because the hypergeometric func-

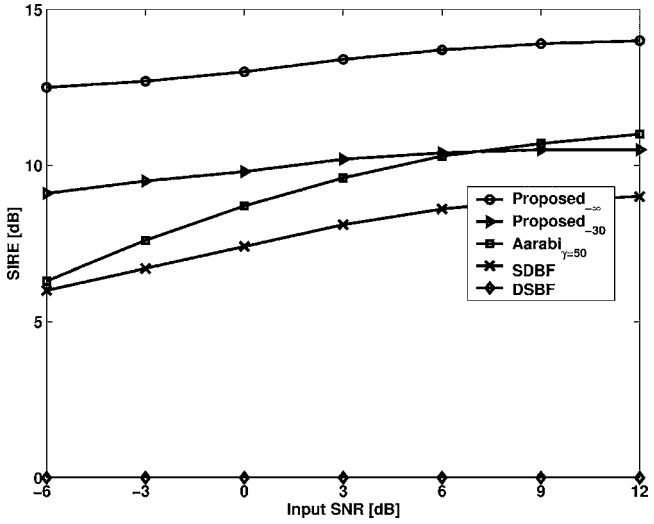


FIG. 18. Canceling of the interference for the driver input (male speaker) for different input SNRs. The input SIRs were kept constant.

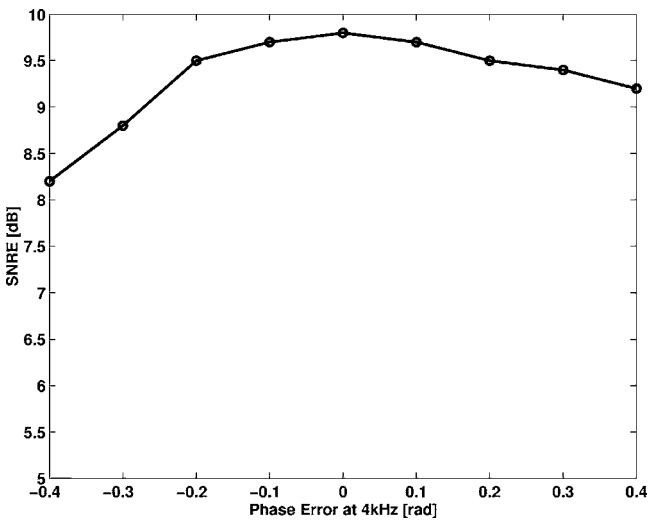


FIG. 20. Susceptibility of the gain function for erroneous phase inputs shown for the driver's signal (male speaker) in terms of the SNRE.

tion used in the gain function is very calculation-intensive, a simplified approximation has to be developed. For real-time implementations, however, it would also be possible to store precalculated values of the full gain function in a table for later use. The development of a gain function which depends on the cross spectrum between more than two microphones could also be an interesting extension, and the use of higher order spectra (Nikias and Petropulu, 1993) might have some potential. Using the new approach in other technical fields like sonar and radar could also open new possibilities for signal extraction.

ACKNOWLEDGMENT

The authors would like to thank the company Ehmann & Partner GmbH, Gundelsheim, Germany for the support with microphones.

APPENDIX A: PDF OF THE PHASE

The direction of arrival for a diffuse noise field has the same probability for any direction. For a sphere with radius 1 and spherical coordinates, $r=1$, $\theta=0,\ldots\pi$, $\phi=0,\ldots,2\pi$, the differential surface element—with $d\phi$ already integrated over $\phi=0,\ldots,2\pi$—is given by

$$dA = r^2 2\pi \sin(\theta) d\theta = 2\pi \sin(\theta) d\theta. \tag{A1}$$

The differential element of the distribution of θ is then given by

$$dP(\theta) = \tfrac{1}{2}\sin(\theta) d\theta, \tag{A2}$$

where Eq. (A1) was divided by the surface area of the unit sphere. Since $p(\theta)=dP(\theta)/d\theta$, the pdf for the angle θ is

$$p(\theta) = \tfrac{1}{2}\sin(\theta). \tag{A3}$$

The phase angle ζ of the cross spectrum is calculated by using the transport time from one microphone to the other (distance d, frequency f, speed of sound c)

$$\zeta = 2\pi \frac{fd}{c} \cos(\theta). \tag{A4}$$

The pdf of ζ is then given using Eq. (5-5) from Papoulis (1991):

$$p(\zeta) = \frac{p(\theta)}{\left|\frac{d\zeta}{d\theta}\right|} = \frac{\frac{1}{2}\sin(\theta)}{2\pi \frac{fd}{c} \sin(\theta)} = \frac{1}{4\pi} \frac{c}{fd}, \tag{A5}$$

and Eq. (5) follows.

APPENDIX B: MATCHING THE NOISE PDFS

The purpose of this appendix is to find the best match of the noise described by the exact pdf Eq. (4) and Eq. (5) using the simplified noise model Eq. (6) by estimating the value of the shifting parameter κ. Because this calculation is not possible in closed form, a numerical solution needs to be found. This can be achieved by multiplying the two pdfs of the exact noise model and the simplified noise model and integrating the resulting function over the entire real and imaginary axis for different values of κ. The value of κ resulting in the highest value of this integration is then the best solution. However, using some straightforward calculations it can be found that the result of κ is independent from speech content.

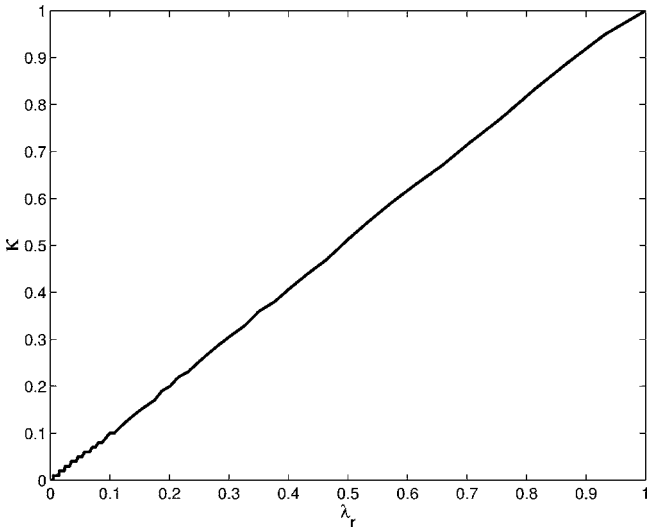


FIG. 21. Optimal value of κ vs the noise variance λ_r.

Using numerical simulations it was found that the relationship between κ and λ_r is approximately linear namely $\kappa \approx a \cdot \lambda_r$ (see Fig. 21 for the case of $\zeta \approx 0$). Figure 22 shows the values of a versus the maximum value of $\zeta_{\max} = 2\pi(fd/c)$, which is the maximum value of ζ in Eq. (5). As $\zeta_{\max} \propto f$ this represents the frequency dependency.

To get good estimates for κ, the value of λ_r needs to be estimated from the signals. As throughout the paper only relatively small microphone distances were used, only the case for $\zeta \approx 0$ is now evaluated: The mean of Eq. (3) with $\zeta=0$ is given by Eq. (4) alone. Using Eq. 3.371 from Gradstein and Ryshik (1981) and some straightforward calculations results in the mean of R namely

$$\mathbf{E}\{R\} = \lambda_r. \tag{B1}$$

The variance is then given by $\mathbf{E}\{(R-\lambda_r)^2\}$ which is—using Eq. 3.351.3 from Gradstein and Ryshik (1981)—resulting in the variance of R given by

$$\sigma_R^2 = 2 \cdot \lambda_r^2. \tag{B2}$$

As σ_R^2 can be approximately measured using the signals by the variance of the real part of the noise cross spectrum $\sigma_R^2 \approx \lambda_{\mathrm{Re}}$, it follows by inserting into Eq. (B2) and solving for λ_r:

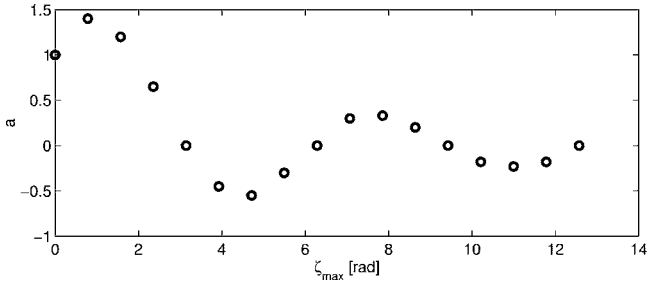


FIG. 22. Different values for the factor a vs the maximum expected noise phase ζ.

$$\lambda_r \approx \sqrt{\frac{\lambda_{\mathrm{Re}}}{2}}. \tag{B3}$$

APPENDIX C: DERIVATION OF THE MMSE GAIN FUNCTION

By using Eqs. (7) and (6) in Eq. (2), the MMSE estimator is

$$\begin{aligned}\hat{S} &= \frac{\int_0^\infty S \exp\left(-\frac{1}{2\lambda_{\mathrm{Re}}}|X\cos(\psi) - S\cos(\varphi_0) - \kappa|^2\right) \exp\left(-\frac{1}{2\lambda_{\mathrm{Im}}}|X\sin(\psi) - S\sin(\varphi_0)|^2\right) \exp\left(-\frac{1}{2\lambda_S}S^2\right) dS}{\int_0^\infty \exp\left(-\frac{1}{2\lambda_{\mathrm{Re}}}|X\cos(\psi) - S\cos(\varphi_0) - \kappa|^2\right) \exp\left(-\frac{1}{2\lambda_{\mathrm{Im}}}|X\sin(\psi) - S\sin(\varphi_0)|^2\right) \exp\left(-\frac{1}{2\lambda_S}S^2\right) dS} \\ &= \frac{\int_0^\infty S \exp\left(-\frac{1}{2\lambda}S^2 + \left(\frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa + \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)S\right) dS}{\int_0^\infty \exp\left(-\frac{1}{2\lambda}S^2 + \left(\frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa + \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)S\right) dS},\end{aligned} \tag{C1}$$

with

$$\frac{1}{\lambda} = \frac{\cos^2(\varphi_0)}{\lambda_{\mathrm{Re}}} + \frac{\sin^2(\varphi_0)}{\lambda_{\mathrm{Im}}} + \frac{1}{\lambda_S}. \tag{C2}$$

Using Eq. 3.462.1 from (Gradstein and Ryshik, 1981),

$$\int_0^\infty x^{\nu-1} e^{-\beta x^2 - \gamma x} dx = (2\beta)^{-\nu/2}\Gamma(\nu)\exp\left(\frac{\gamma^2}{8\beta}\right) D_{-\nu}\left(\frac{\gamma}{\sqrt{2\beta}}\right),$$

$$[\mathrm{Re}\,\beta > 0, \mathrm{Re}\,\nu > 0],$$

the result for the estimate of the speech amplitude $\hat{S}$ is

$$\begin{aligned}\hat{S} &= \frac{\lambda\Gamma(2)\exp\left(\frac{\lambda}{4}\left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa - \frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)^2\right) D_{-2}\left(\sqrt{\lambda}\left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa - \frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)\right)}{\sqrt{\lambda}\Gamma(1)\exp\left(\frac{\lambda}{4}\left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa - \frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)^2\right) D_{-1}\left(\sqrt{\lambda}\left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa - \frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)\right)} \\ &= \sqrt{\lambda}\frac{D_{-2}\left(\sqrt{\lambda}\left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa - \frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)\right)}{D_{-1}\left(\sqrt{\lambda}\left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}}\kappa - \frac{X}{\lambda_{\mathrm{Re}}}\cos(\psi)\cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}}\sin(\psi)\sin(\varphi_0)\right)\right)},\end{aligned} \tag{C3}$$

where $\Gamma(\cdot)$ is the gamma function and $D_p(z)$ is the parabolic cylinder function (Gradstein and Ryshik, 1981, Eq. 9.240). This is given by

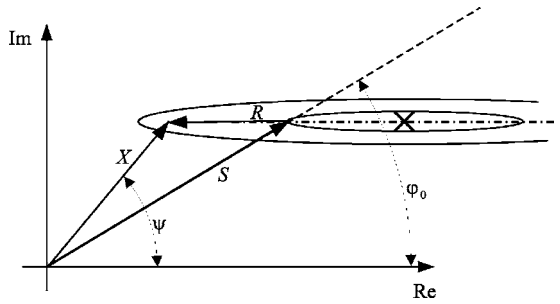


FIG. 23. Maximum gain value at angles greater than φ_0.

$$D_p(z) = 2^{p/2} e^{-z^2/4} \left(\frac{\sqrt{\pi}}{\Gamma\left(\frac{1-p}{2}\right)} \Phi\left(-\frac{p}{2}, \frac{1}{2}; \frac{z^2}{2}\right) - \frac{\sqrt{2\pi} z}{\Gamma\left(-\frac{p}{2}\right)} \Phi\left(\frac{1-p}{2}; \frac{3}{2}; \frac{z^2}{2}\right) \right), \tag{C4}$$

where $\Phi(\alpha, \beta; z)$ is the confluent hypergeometric function (Gradstein and Ryshik, 1981, Eq. 9.210):

$$\Phi(\alpha, \beta; z) = 1 + \frac{\alpha}{\beta} \frac{z}{1!} + \frac{\alpha(\alpha+1)}{\beta(\beta+1)} \frac{z^2}{2!} + \frac{\alpha(\alpha+1)(\alpha+2)}{\beta(\beta+1)(\beta+2)} \frac{z^3}{3!} + \ldots . \tag{C5}$$

Using the definitions

$$\upsilon = \frac{\xi}{\left(\cos^2(\varphi_0) \sqrt{\frac{\lambda_{\mathrm{Im}}}{\lambda_{\mathrm{Re}}}} + \sin^2(\varphi_0) \sqrt{\frac{\lambda_{\mathrm{Re}}}{\lambda_{\mathrm{Im}}}} \right) \xi + 1} \gamma, \tag{C6}$$

$$\xi = \frac{\lambda_S}{\sqrt{\lambda_{\mathrm{Re}} \lambda_{\mathrm{Im}}}}, \tag{C7}$$

and

$$\gamma = \frac{X^2}{\sqrt{\lambda_{\mathrm{Re}} \lambda_{\mathrm{Im}}}}, \tag{C8}$$

which are interpreted [following (Ephraim and Malah (1984) and McAulay and Malpass (1980)] as the *a priori* SNR and the *a posteriori* SNR, as well as pooling the argument of $D_p(\cdot)$ in Eq. (C3),

$$\Theta = \sqrt{\lambda} \left(\frac{\cos(\varphi_0)}{\lambda_{\mathrm{Re}}} \kappa - \frac{X}{\lambda_{\mathrm{Re}}} \cos(\psi) \cos(\varphi_0) - \frac{X}{\lambda_{\mathrm{Im}}} \sin(\psi) \sin(\varphi_0) \right), \tag{C9}$$

the MMSE-gain function Eq. (8) results from Eq. (C3).

APPENDIX D: ANGLE OF MAXIMUM GAIN VALUE

One might have noticed (see Figs. 7–9), that the greatest values of the gain function $\hat{G}$ occur at angles ψ exceeding φ_0. This behavior will now be explained using Fig. 23, where the ellipses show equivalent values of the error's pdf.

The observed cross spectrum is $Xe^{j\psi}$, where $\psi > \varphi_0$. Using the model given in Sec. IV, this is, most likely, a result of the shown summation of the speech-component $Se^{j\varphi_0}$ and the noise-component $Re^{j\zeta}$ (the angle ζ is not shown for clarity). Hence the amplitude of the speech-component S is greater than the observed amplitude X, a gain value $\hat{G} > 1$ results.

[1]The evaluation of the gain function using Eqs. (C6)–(C8) is based on the analytical result. Otherwise the solution for the MMSE estimator has to be calculated numerically using Eqs. (4) and (5).

[2]This is equivalent to Eq. (A4), where ζ was replaced by φ_0. The direction of incidence is represented by θ.

Aarabi, P., and Shi, G. (**2004**). "Phase-based dual-microphone robust speech enhancement," IEEE Trans. Syst., Man, Cybern., Part B: Cybern. **34**, 1763–1773.

Boll, S. F. (**1979**). "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-27**, 113–120.

Brandstein, M., and Ward, D., eds. (**2001**). *Microphone Arrays*, (Springer, Berlin).

Cohen, I. (**2003**). "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," IEEE Trans. Speech Audio Process. **11**, 466–475.

Dahl, M., and Claesson, I. (**1999**). "Acoustic noise and echo cancelling with microphone array," IEEE Trans. Veh. Technol. **48**, 1518–1526.

Doblinger, G. (**1995**). "Computationally efficient speech enhancement by spectral minima tracking in subbands," EURASIP: Proceedings of EUROSPEECH, Madrid, Spain, Vol. **2**, pp. 1513–1516.

Dörbecker, M. (**1997**). "Small microphone arrays with optimized directivity for speech enhancement," EURASIP: Proceedings of EUROSPEECH, Rhodes, Greece.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-32**, 1109–1121.

Gilbert, E. N., and Morgan, S. P. (**1955**). "Optimum design of directive antenna arrays subject to random variations," Bell Syst. Tech. J. **34**, 637–663.

Gradstein, I. S., and Ryshik, I. M. (**1981**). *Tables of Series, Products, and Integrals* (Deutsch, Thun).

Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (**2003**). "Evaluation of array-processing algorithms for a headband hearing aid," J. Acoust. Soc. Am. **113**, 1646–1657.

Griffiths, L. J., and Jim, C. W. (**1982**). "An alternative approach to linear constrained adaptive beamforming," IEEE Trans. Antennas Propag. **AP-30**, 27–34.

Gustafsson, H., Nordholm, S. E., and Claesson, I. (**2001**). "Spectral subtraction using reduced delay convolution and adaptive averaging," IEEE Trans. Speech Audio Process. **9**, 799–807.

Gustafsson, S. (**1999**). "Enhancement of audio signals by combined acoustic echo cancellation and noise reduction," Ph.D. thesis, RWTH Aachen, Aachen, Germany.

Knecht, W. G., Schenkel, M. E., and Moschytz, G. S. (**1995**). "Neural network filters for speech enhancement," IEEE Trans. Speech Audio Process. **3**, 433–438.

Lotter, T., Benien, C., and Vary, P. (**2003**). "Multichannel direction-independent speech enhancement using spectral amplitude estimation," EURASIP Journal on Applied Signal Processing EURASIP **2003**, pp. 1147–1156.

Marro, C., Mahieux, Y., and Simmer, K. U. (**1998**). "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," IEEE Trans. Speech Audio Process. **6**, 240–259.

Martin, R. (**2001**). "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech Audio Process. **9**, 504–512.

McAulay, R. J., and Malpass, M. L. (**1980**). "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-28**, 137–145.

Meyer, J. (**2001**). "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**, pp. 185–193.
Nikias, C. L., and Petropulu, A. P. (**1993**). *Higher-Order Spectra Analysis* (Prentice Hall, Upper Saddle River, NJ).
Papoulis, A. (**1991**). *Probability, Random Variables, and Stochastic Processes* 3rd ed. (WCB/McGraw-Hill, Boston).
Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice Hall, Englewood Cliffs, NJ).
Seltzer, M. L., and Raj, B. (**2003**). "Speech-recognizer-based filter optimization for microphone array processing," IEEE Signal Process. Lett. **10**, 69–71.
Soon, I. Y., and Koh, S. N. (**2000**). "Low distortion speech enhancement," IEE Proc. Vision Image Signal Process. **147**, 247–253.
Vanden Berghe, J., and Wouters, J. (**1998**). "An adaptive noise canceller for hearing aids using two nearby microphones," J. Acoust. Soc. Am. **103**, 3621–3626.

An inversion of Freedman's "image pulse" model in air

Janos Tsakiris[a] and Phillip McKerrow
Intelligent Robotics Laboratory, School of Information Technology and Computer Science, University of Wollongong, Wollongong, NSW, 2522, Australia

(Received 22 June 2005; revised 17 November 2005; accepted 19 November 2005)

In recent years, research into object classification based on ultrasonic sensing has shown that sonar is a rich source of data suitable for robust classification for specific classes of objects. However, these systems lack generality, precision, and are often slow as they need to collect and integrate multiple scans. The objective was the development of more efficient techniques for ultrasonic based object recognition through an inversion of the extended form of Freedman's "image pulse" model. Earlier work extended this model, originally developed for a fluid medium with coincident transmitter/receiver configurations, to noncoincident configurations in air. This extended model was inverted, producing one that would calculate the geometry of a scattering body from an analysis of echoes received after insonification of the body with ultrasonic pulses. Quantitative verification of this model with various scattering bodies proved elusive, with low correlation between experiment and theory due to matrix instability and difficulties in obtaining data of sufficient accuracy. However, qualitative trends in the data indicate the model is essentially correct, though very sensitive to measurement precision and media characteristics, and there is reason to believe that further work under more controlled laboratory conditions and/or a different medium would verify the model's validity quantitatively.
[DOI: 10.1121/1.2151791]

PACS number(s): 43.60.Lq, 43.60.Pt, 43.35.Ae [EJS] Pages: 965–975

I. INTRODUCTION

The viability of autonomous robotics depends heavily on the development of robust methods for sensing the physical environment and their use as a basis for reliable navigation. Vision based approaches are popular at present, but ultrasonic sensing has been used for some time in parallel to these and has demonstrated itself to be a rich source of data for robust object classification for specific classes of objects.[1] There are drawbacks however, namely, in the variety of objects that can be detected, the accuracy to which the physical parameters of these objects can be determined, and the time taken to insonify them. The intent was to develop a method of sensing that would address some of these drawbacks via the development of an inverse model enabling a fast and accurate determination of the physical characteristics of an object from a measurement of its echo after insonification.

The inversion was based on Freedman's "image pulse" model of acoustic propagation.[2] Although originally designed for use underwater, work was completed previously in adapting this model for an atmospheric environment.[3] This paper addresses the work that follows on from this, covering the development of the inverse model, the methodology by which the model was tested and an analysis of the limitations of these testing procedures.

A quantitative verification of the inverse model proved elusive due to the coupling of an ill conditioned matrix with data of insufficient accuracy, producing estimates that had a low correlation with the measured data. The experimental results presented here confirm this conclusion but trends in the measured data indicate that the model is essentially correct though very sensitive to measurement precision and media characteristics.

II. DEVELOPMENT OF THE INVERSE MODEL

The central tenet of Freedman's model is that an insonified object generates echoes at those points on its surface where there is a discontinuity in any of the derivatives of the solid angle subtended at the receiver with respect to range. To illustrate this, Fig. 1 shows the echoes generated from a sphere with a radius of a, the point nearest to the transmitter located at r_1 and its equator located at r_2. The extended form of Freedman's model is given in Eq. (1) providing the time-varying voltage E_g measured at a receiver located a distance r_g from the source of an echo of frequency f

$$E_g = A(f)\frac{GV\beta_P(\phi)\beta_H(\phi)}{r_g^2}\exp(-2r_g\alpha) \times\left[D(0)\sin\theta - \frac{D(1)}{2k}\cos\theta - \frac{D(2)}{4k^2}\sin\theta\right], \quad (1)$$

where $A(f)$ is the transfer function of the transducer and receiver at frequency f, G is the gain, or amplification, applied to the received signal, V is the voltage applied to the transducer, $\beta_P(\phi)$ is the frequency-dependent directivity of the transducer, $\beta_H(\phi)$ is the frequency-dependent directivity of the receiver, r_g is the range to the source of the echo, α is the frequency-dependent atmospheric absorption coefficient, $D(n)$ is the discontinuity equations for the nth order derivatives, k is the frequency dependent wave number, and $\theta = \omega t - 2kr$.

[a]Electronic mail: janos.tsakiris@gmail.com

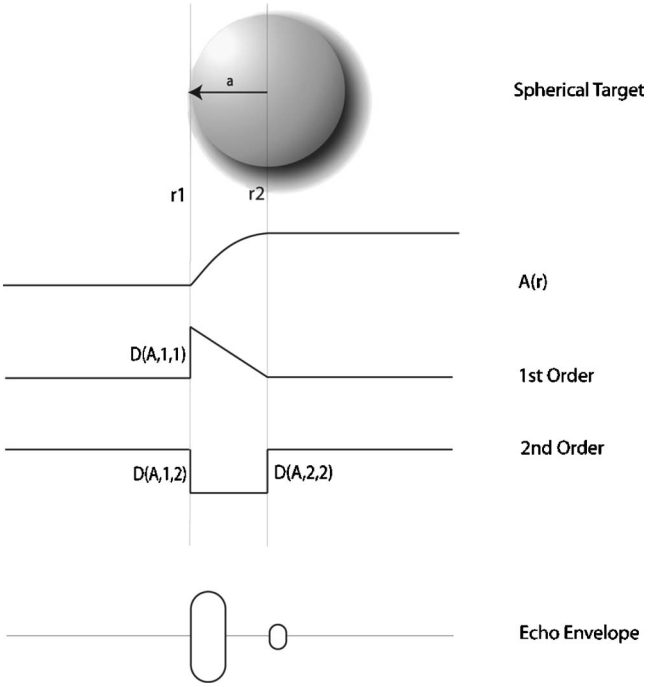


FIG. 1. Echoes from a sphere of radius a as predicted by Freedman's model.

The model is founded upon several assumptions.[2] The transmitter/receiver and the scattering body must be stationary to eliminate Doppler shift, variable ranges and the fluid dynamics of a body in transit through a fluid. The scatterer is assumed to be rigid (no internal wave penetrations) and convex (no multiple reflections). Further, it is assumed that the amplitudes of the waves are too small to cause any motion in the body itself (to prevent resonance phenomena) and the wavelength used is much smaller than the dimensions and radius of curvature of the scattering body (a requirement of Kirchoff's approximation, used in the derivation of the forward model). Also, it is assumed that the surface of the scatterer is smooth, to prevent diffuse reflection. The fluid medium that encompasses the transmitter, receiver, and scattering body is assumed to be nondissipative as well as both homogeneous and isotropic. The scattering body is also assumed to be in the Fraunhofer region (i.e., far field) of the transmitter and receiver. Last, an assumption specific to the extended form of the forward model is that the scatterer must have a smoothly varying surface, which is to say no discontinuities in $D(0)$ across its surface.

For reasons that will become clear, it was prudent to redefine some of the elements of Eq. (1), with emphasis on the frequency dependence, or otherwise, of each of the components. By combining several components into the function $K(f)$, one gets

$$K(f) = \frac{A(f)GV\beta_P(\phi)\beta_H(\phi)}{r_g^2} \exp(-2r_g\alpha). \tag{2}$$

Equation (2), as its formulated above, implies $K(f)$ is dependent solely upon frequency. This is not true in the general case, however it will be demonstrated that when a stationary object is insonified with several different frequencies within the span of a couple of minutes, $K(f)$ is effectively dependent upon frequency alone.

First, during the chirp process the sensitivity of the transducer is dependent upon both (a) the tension in the Mylar film and (b) the frequency of the signal. The tension in the Mylar reduces negligibly over the duration of a couple of minutes, especially since the bias is turned off between measurements as a precaution to ensure this and thus it is effectively constant, leaving the transducer dependent only upon frequency. As the sensitivity of the microphone is also dependent upon frequency alone, it follows that the transfer function $A(f)$ for the transducer and receiver is dependent solely upon frequency, in this case.

The directivities $\beta_P(\phi)$ and $\beta_H(\phi)$ of the transducer and receiver, respectively, are functions of (a) the frequency of the wave, (b) its speed c, and (c) the angular bearing ϕ of the scattering object to their axes. The speed of sound c is dependent upon the environmental variables of temperature, pressure, and humidity, each of which do not change sufficiently in the span of two minutes to produce a measurable change in the speed of sound, thus rendering the speed of sound a constant during this period. The bearing ϕ is also constant in this scenario as the scattering object is stationary. Thus $\beta_P(\phi)$ and $\beta_H(\phi)$ are dependent upon frequency alone, in this case.

Last, the atmospheric absorption coefficient α, having dependence upon temperature, pressure, and humidity, in addition to frequency, remains a function of frequency alone in this case, as it has been established that these three environmental variables will remain virtually unchanged over a two minute span. Given that the remaining factors of G, V, and r_g were established in Eq. (1) as being constants, it follows from this analysis that during the measurement cycle, the function $K(f)$ has dependence on frequency alone. Reformulating Eq. (1) for this case

$$E_g(f) = K(f)\left[D(0)\sin\theta - \frac{D(1)}{2k}\cos\theta - \frac{D(2)}{4k^2}\sin\theta\right]$$
$$= Y_1(f)D(0) + Y_2(f)D(1) + Y_3(f)D(2), \tag{3}$$

where, $Y_1(f) = K(f)\sin\theta$, $Y_2(f) = -K(f)/2k\cos\theta$, $Y_3(f) = -K(f)/4k^2\sin\theta$, and $\theta = wt - 2kr_g = 2\pi ft - 2kr_g$.

Equation (3) still varies in the time domain by virtue of θ. In order for the inversion to proceed however, the variance of time t must be nullified to ensure that Eq. (3) is dependent solely upon one variable, the frequency f. For now, it will be assumed that t is constant, the details of how this manifests itself practically being explored later. Using three distinct frequencies, the measured voltages at the receiver at time t for each frequency can be given by

$$E_g(f_1) = Y_1(f_1)D(0) + Y_2(f_1)D(1) + Y_3(f_1)D(2), \tag{4}$$

$$E_g(f_2) = Y_1(f_2)D(0) + Y_2(f_2)D(1) + Y_3(f_2)D(2), \tag{5}$$

$$E_g(f_3) = Y_1(f_3)D(0) + Y_2(f_3)D(1) + Y_3(f_3)D(2). \tag{6}$$

In matrix form

$$\begin{bmatrix} E_g(f_1) \\ E_g(f_2) \\ E_g(f_3) \end{bmatrix} = \begin{bmatrix} Y_1(f_1) & Y_2(f_1) & Y_3(f_1) \\ Y_1(f_2) & Y_2(f_2) & Y_3(f_2) \\ Y_1(f_3) & Y_2(f_3) & Y_3(f_3) \end{bmatrix} \begin{bmatrix} D(0) \\ D(1) \\ D(2) \end{bmatrix} \tag{7}$$

and inverted

$$\begin{bmatrix} D(0) \\ D(1) \\ D(2) \end{bmatrix} = \begin{bmatrix} Y_1(f_1) & Y_2(f_1) & Y_3(f_1) \\ Y_1(f_2) & Y_2(f_2) & Y_3(f_2) \\ Y_1(f_3) & Y_2(f_3) & Y_3(f_3) \end{bmatrix}^{-1} \begin{bmatrix} E_g(f_1) \\ E_g(f_2) \\ E_g(f_3) \end{bmatrix}. \tag{8}$$

Thus Eq. (8) should enable the geometric properties of $D(0)$, $D(1)$, and $D(2)$ of the insonified object to be deduced from the measured echoes $E_g(f_n)$ at the receiver and a knowledge of $Y_x(f_y)$. The definitions of $Y_x(f)$ in Eq. (3) show that they are dependent upon the wave number k and $K(f)$. As the frequency f is known for each of the frequencies used and the speed of sound c is easily calculated from knowledge of the measured environmental conditions, it follows that the wave number k is easily calculated for each frequency using $k=2\pi f/c$.

To calculate $K(f)$ for each frequency, values for each of the terms in Eq. (2) need to be obtained. The drawback of this approach is however that errors in each of these accumulate when combined into Eq. (2). Another approach, both faster and more accurate, is to calibrate the model with an object of precisely known dimensions and determine $K(f)$ directly through measurement. Typically, a calibration object that has all $D(n) = 0$, except for one, would be the ideal choice, such as a paraboloid for instance, having all discontinuities equal to zero except $D(1)$. In this case, Eq. (3) would be formulated as

$$E_g = -K(f)\frac{D(1)}{2k}\cos\theta \tag{9}$$

making the calculation of $K(f)$ at each frequency relatively trivial, given that all the other components in the equation are either easily calculated (i.e., k, θ), easily measured (i.e., E_g) or already known [i.e., $D(1)$]. Time t can be kept constant by sampling the echo at the same time index for each of the three frequencies. Once $K(f)$ is known for each frequency, $Y_x(f)$ for each is easily calculated. These are then substituted into Eq. (8), thereby determining the discontinuities in the scattering object.

As the inverse model is based on the forward model given in Eq. (1), the constraints upon that model must also apply here as well. One such constraint is that the scatterer has a smoothly varying surface. Thus it can be assumed that $D(0)=0$ and Eq. (3) is modified accordingly

$$E_g(f) = Y_2(f)D(1) + Y_3(f)D(2), \tag{10}$$

where $Y_2(f)=-K(f)/2k\cos\theta$ and $Y_3(f)=-K(f)/4k^2\sin\theta$.

Thus, the inverse model for a scatterer with a smoothly varying surface [i.e., $D(0)=0$] can be given by

$$\begin{bmatrix} D(1) \\ D(2) \end{bmatrix} = \begin{bmatrix} Y_2(f_1) & Y_3(f_1) \\ Y_2(f_2) & Y_3(f_2) \end{bmatrix}^{-1} \begin{bmatrix} E_g(f_1) \\ E_g(f_2) \end{bmatrix}. \tag{11}$$

III. VISUALIZATION OF THE SCATTERING BODY

Knowledge of $D(n)$ only hints at the shape of the insonified object. To fully represent its geometry, a visual model of the scatterer needs to be built using numerical methods. For the sake of generality, the following will examine how this can be done for all cases, not just for $D(0)=0$.

First, $D(0)$, $D(1)$, and $D(2)$ are obtained for each range along the scatterer for which there is an echo. At the front face of the scatterer at a range r_{0-} the discontinuity equations are

$$D(A,0,0) = A(r_{0-}) - A(r_{0+}), \tag{12}$$

$$D(A,0,1) = \frac{dA}{dr}(r_{0-}) - \frac{dA}{dr}(r_{0+}), \tag{13}$$

$$D(A,0,2) = \frac{d^2A}{dr^2}(r_{0-}) - \frac{d^2A}{dr^2}(r_{0+}). \tag{14}$$

$d^nA/dr^n(r_{0-})$ are all zero, as r_{0-} is the range immediately prior to the front face where there is no variation in the cross-sectional area (being empty space). Hence

$$\text{Area at } r_{0-} = -D(A,0,0), \tag{15}$$

$$\frac{dA}{dr} \text{ at } r_{0-} = -D(A,0,1), \tag{16}$$

$$\frac{d^2A}{dr^2} \text{ at } r_{0-} = -D(A,0,2). \tag{17}$$

If it is assumed that the second-order derivatives remains constant over the range r_{0+} to r_{m-} [i.e., $d^2A/dr^2(r_{m-}) = d^2A/dr^2(r_{0+})$] where r_m is the range at which the next echo originates, then the following relations can be applied at each successive point within this range,

$$A_{x+1} = A_x + \frac{dA_x}{dr}(r_{x+1} - r_x), \tag{18}$$

$$\frac{dA_{x+1}}{dr} = \frac{dA_x}{dr} + \frac{d^2A_x}{dr^2}(r_{x+1} - r_x). \tag{19}$$

Even if the assumption of a constant second-order derivative does not hold, the result is at least an approximation of the object's geometry, the accuracy of this projection depending upon how much variance the second-order derivative undergoes over the extent of the scatterer. Once $A(r_{m-})$ and $dA/dr(r_{m-})$ have been determined, the area and higher-order rates of change at r_m can be calculated using the following:

$$\text{Area at } r_m = A(r_{m-}) - D(A,m,0), \tag{20}$$

$$\frac{dA}{dr} \text{ at } r_m = \frac{dA}{dr}(r_{m-}) - D(A,m,1), \tag{21}$$

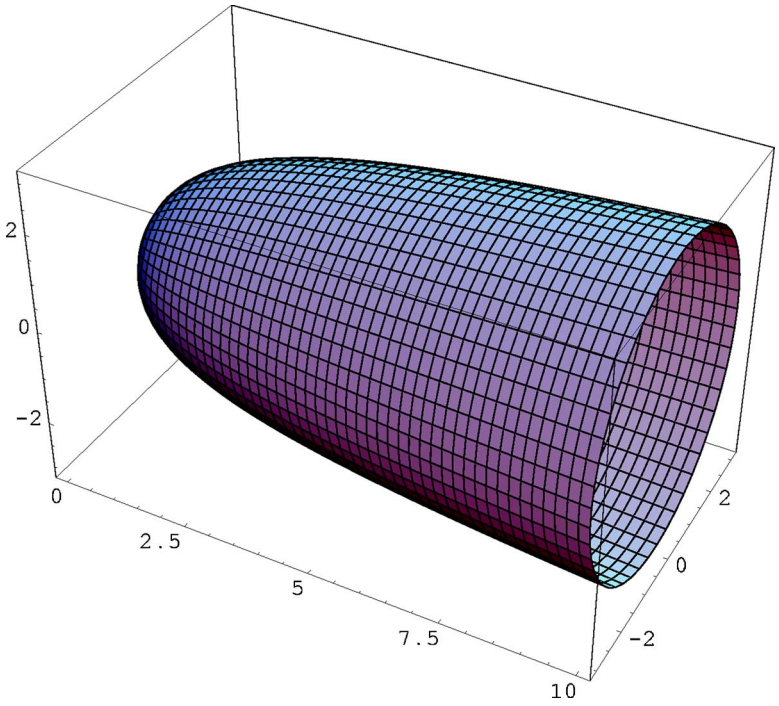


FIG. 2. Color (online) Representation of an aluminium object with the geometry governed by Eq. (24).

$$\frac{d^2A}{dr^2} \text{ at } r_m = \frac{d^2A}{dr^2}(r_{m-}) - D(A,m,2). \tag{22}$$

This process is continued for all ranges at which echoes originate and thereby the shape of the scatterer, or at least an approximation thereof, is reconstructed.

IV. GEOMETRIC ANALYSIS OF SCATTERING BODIES

To test the validity of the inverse model, a series of objects for use as scattering bodies with physical characteristics satisfying the assumptions of the model needed to be obtained. Specifically, rigid objects with smoothly varying surfaces and separations between their discontinuity points such that no overlap of echoes would occur. To accomplish this, a series of aluminium objects were machined with geometrically simple shapes. The shapes chosen were arbitrary, provided that they represent a broad spectrum of possible objects that would fit the description of being smoothly varying. These were machined to precise dimensions, imperfections in the final products from this process being under 0.2 mm, well under the limit required by the Raleigh criterion.

The first object in the set had an exponential geometry with a length of 10 cm and a cross-sectional area varying from 0 cm^2 at its vertex to 30.2 cm^2 at its base, its geometry being governed by Eq. (23)

$$x(r) = \exp\left(\frac{r^2}{4}\right) - 1, \tag{23}$$

where $x(r)$= the length (cm) along the axis, measured from the vertex and r= the radius of the cross-sectional area of the object.

As Freedman's model uses the variation of cross-sectional area with range, it is appropriate to express Eq. (23) with the cross-sectional radius as a function of range, rather than the converse, producing Eq. (24),

$$r(x) = \sqrt{4\ln(x+1)}. \tag{24}$$

The geometry of this object is represented in Fig. 2, where the axes are in units of centimeters. Despite the hollow appearance of the rendered image in this figure, the objects used were solid throughout. A more useful visual representation of the structure of the object is shown in Fig. 3, a plot of the variation of the cross-sectional area $A(x)$ with respect to range and governed by Eq. (25),

$$A(x) = 4\pi\ln(x+1). \tag{25}$$

The derivative of Eq. (25) with respect to range yields the following:

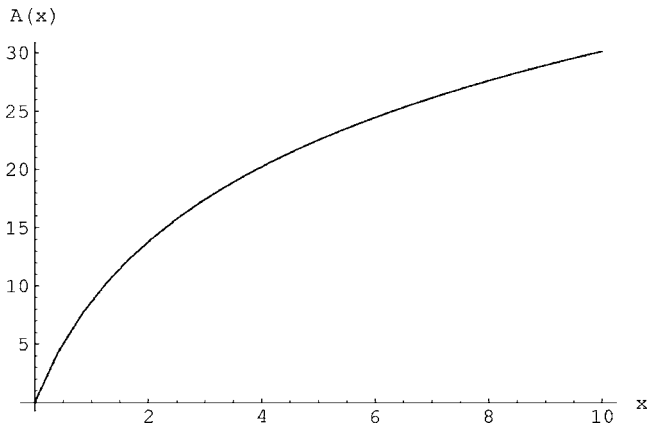


FIG. 3. Plot of the cross-sectional area (cm^2) with respect to range (cm) for the object in Fig. 2.

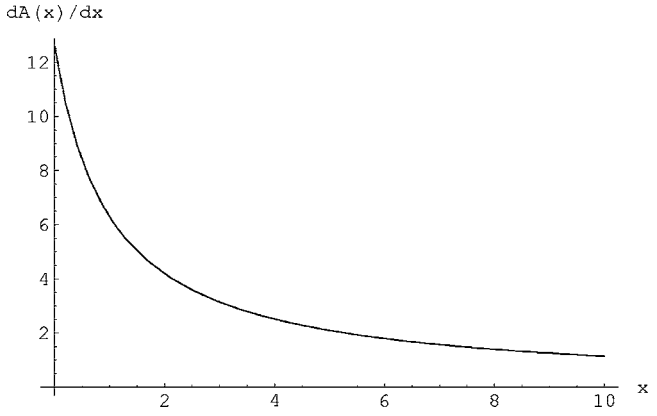


FIG. 4. Derivative of the cross-sectional area with respect to range (cm) of the object in Fig. 2.

$$\frac{dA(x)}{dx} = \frac{4\pi}{1+x}, \tag{26}$$

a plot for which is given in Fig. 4. Taking the derivative of Eq. (26) provides the second order derivative in Eq. (27),

$$\frac{d^2A(x)}{dx^2} = \frac{-4\pi}{(1+x)^2} \tag{27}$$

represented in Fig. 5. The derivatives now calculated, it is clear from Figs. 2–4 that they are all *continuous* functions, having no step discontinuities in Eqs. (25)–(27) from 0 to 10 cm and hence no echoes. At $x=0$ cm however, there is a transition between the free space prior to the object, where the cross-sectional area (and its higher-order derivatives) are zero, and the start of the object itself. To see whether the transition across these two cross-sectional profiles results in step discontinuities (and echoes) at this point, the plots shown in Figs. 2–4 can be examined again.

Figure 3 reveals a cross-sectional area of zero at $x=0$ cm. As this is equal to that of the free space region, there is no discontinuity in the cross-sectional area here point and hence no echo. Expressed quantitatively,

$$D(A_w,g,n) = \frac{d^nA_w}{dr^n}(r_{g-}) - \frac{d^nA_w}{dr^n}(r_{g+}), \tag{28}$$

where r_{g-} and r_{g+} are the ranges infinitesimally prior to and beyond the discontinuity. In this case, $n=0$ and $g=0$ cm. Thus,

$$D(A_w,0,0) = A_w(r_{0-}) - A_w(r_{0+}). \tag{29}$$

Since $A_w(r_{0-})$ represents the cross-sectional area in the free space immediately prior to the start of the object, it is necessarily equal to zero. As for $A_w(r_{0+})$, it can be seen from Fig. 3 that at a point infinitesimally beyond $x=0$ cm, the cross-sectional area remains zero. Since these quantities are equal, Eq. (29) is zero, indicating that there is no discontinuity in the cross-sectional area profile at this point and thus this factor would not contribute to any echoes.

For the higher-order derivatives, Fig. 4 is examined next. At $x=0$ cm, the derivative of the cross-sectional area with respect to range is 4π. Given that it has been established that the free space region prior to the scatterer has a cross-sectional area of zero, at all orders, it follows that the transition at $x=0$ cm from a cross-sectional area of zero to 4π must result in a step discontinuity. Thus,

$$D(A_w,0,1) = \frac{dA_w}{dr}(r_{0-}) - \frac{dA_w}{dr}(r_{0+}) = (0) - (4\pi) = -4\pi. \tag{30}$$

Consequently, at $x=0$ cm, there *is* a contribution to the echo as a result of the discontinuity in the *derivative* of the cross-sectional area with respect to range.

Finally, in Fig. 5 the second-order derivative of cross-sectional area is -4π at the transition point $x=0$ cm. For the reasons outlined, there is therefore a step discontinuity at this point given by

$$D(A_w,0,2) = \frac{d^2A_w}{dr^2}(r_{0-}) - \frac{d^2A_w}{dr^2}(r_{0+}) = 0 - (-4\pi) = 4\pi. \tag{31}$$

Therefore there is also a contribution to the echo originating at $x=0$ cm from the discontinuity in the second-order de-

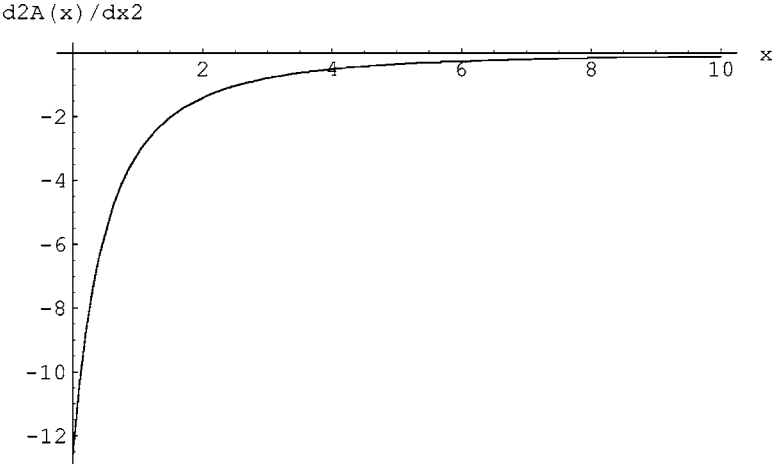


FIG. 5. Second-order derivative of the cross-sectional area with respect to range (cm) of the object in Fig. 2.

TABLE I. Locations and magnitudes for discontinuities on all test objects.

Object	$r(x)$ (cm)	x (cm)	$D(n)$ (cm)	$D(0)$	$D(1)$	$D(2)$
Exponential	$\sqrt{4\ln(x+1)}$	$0 \leqslant x \leqslant 10$	0	0	-4π	4π
Paraboloid A	$\sqrt{12x/5}$	$0 \leqslant x \leqslant 10$	0	0	-2.4π	0
Spheroid-Paraboloid	$\sqrt{20x-x^2}$	$0 \leqslant x \leqslant 5$	0	0	-20π	2π
	$\sqrt{7.05(x+5.64)}$	$5 \leqslant x \leqslant 10$	5	0	2.95π	-2π
Paraboloid B	$\sqrt{4.9x}$	$0 \leqslant x \leqslant 10$	0	0	-4.9π	0
Ellipsoid	$\sqrt{4x(10-x)}$	$0 \leqslant x \leqslant 5$	0	0	-40π	8π
Fourth order	$3.15x^{1/4}$	$0 \leqslant x \leqslant 10$	0	0	$-\infty$	$+\infty$

rivative of the cross-sectional area with respect to range. Despite the magnitude of the discontinuity being the same in both cases (i.e., 4π) the size of the contribution of each component to the total echo will differ however. This is due to Freedman's equation weighting successively higher-order discontinuities several orders of magnitude lower than their predecessors.

Clearly, since this object, as with all the objects, has a sharp edge at the base (the transition range between the insonified and shadow regions) echoes would be expected from that region. However, as such regions do not hold with the assumption of smooth surface variance though, the focus is only on the echoes that originate at those points on the scatterer where the surface is smoothly varying. Since the echoes from the base are separate from the front face echoes in the time domain, there is no cross contamination and the base echoes can be ignored.

The locations and magnitudes of the nth order discontinuities of the cross-sectional areas with respect to range for each of the objects in the set were calculated, the results for which are given in Table I. Special mention needs to be made of the calculated discontinuities on the final object. Despite the first- and second-order discontinuities being of infinite magnitude, in practice the echo must be finite, albeit of large magnitude. The fact that a higher-order discontinuity of infinite magnitude is incorrectly processed by the model as producing an echo of infinite magnitude, with the converse holding true for the inverse model, highlights one of the limitations of the model.

V. EXPERIMENTAL DESIGN, RESULTS AND ANALYSIS

Before testing of the model using these objects began, the frequencies that would be used to insonify them needed to be chosen, as well as the range at which the objects would be placed. Given that the Polaroid transducer has a maximum output at 50 kHz, with a gradual falloff in sensitivity above and below this frequency, it was decided that the chosen frequencies should fall on either side of this peak output. Thus 40 and 60 kHz were selected as they are each close enough to 50 kHz to provide a signal which is not too weak while at the same time sufficiently separated along the spectral band in order that they should produce echoes that are noticeably unique from one another.

A range of 40 cm was selected. Even though this range is borderline for the validity of the Kirchoff approximation, preliminary tests showed that the echoes at ranges beyond this were too weak to reasonably expect to detect contributions from the higher-order discontinuities. As the detectable presence of these higher-order contributions is necessary for the inverse model's viability, the range had to be limited accordingly.

The procedure for obtaining the echoes was not dissimilar to that employed earlier[3] when the validity of the extended forward model was tested. Software settings on the chirp/capture software also mimicked those used during the forward model tests, specifically a ten cycle wave is emitted, the echo sampled at a rate of 1 MHz and the procedure repeated 100 times. This multi-sampling is necessary for the same reasons that it was employed for the earlier tests, to minimize random, short-term fluctuations in the environment.

With the apparatus now prepared, the first object was mounted at 40 cm and the measurements commenced. At the conclusion of the first measurement cycle, the frequency was changed from 40 to 60 kHz and the entire cycle repeated. Once the readings from the object at both frequencies had been obtained, the environmental conditions were recorded, the object was demounted, the next object in the series put in its place, and the process repeated until all six objects had been examined in this manner.

A. Determining $K(f)$

First, the matrix elements of the inverse model must be determined. As stated earlier, in order to calculate the matrix elements $Y_x(f_y)$ of the model, $K(f)$ should be measured by calibrating the model with an object of precisely known dimensions, a paraboloid being the best choice. Of the two paraboloids in the set, Paraboloid B, the dimensions for which are given in Table I, was deemed the most suitable as it has a larger discontinuity which should produce a stronger echo and thus a more accurate result. Using Paraboloid B, with $D(1)=-4.9\pi$, and substituting into Eq. (9),

$$E_g = K(f)\frac{2.45\pi}{k}\cos\theta. \tag{32}$$

At its peak, when $\cos\theta=1$, the voltage has a maximum value given by

$$E_{g_{\mathrm{MAX}}} = K(f)\frac{2.45\pi}{k}, \tag{33}$$

and given that $k=2\pi f/c$, Eq. (33) can be rearranged as

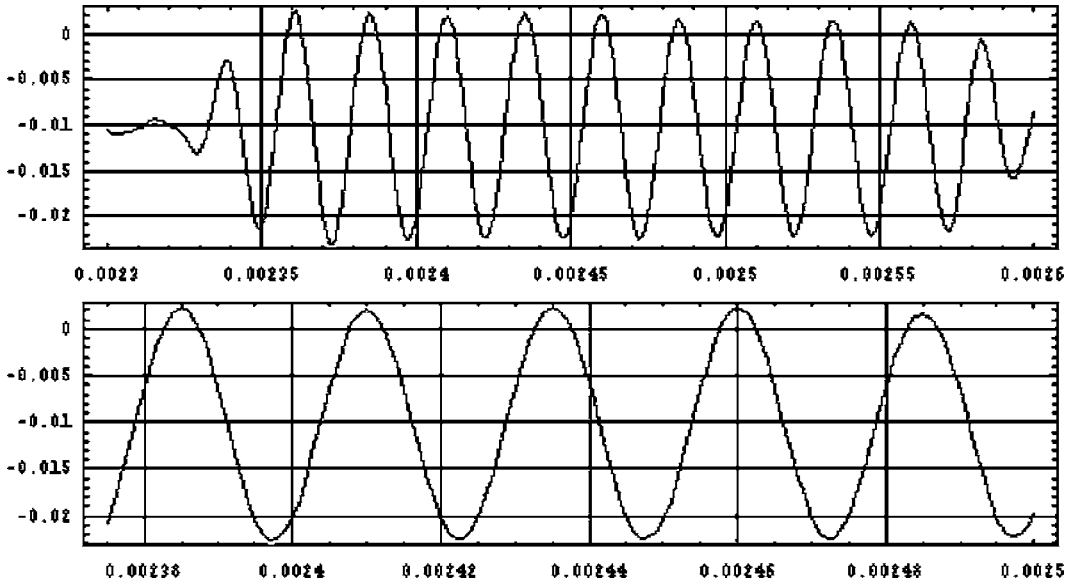


FIG. 6. Original echo data from the exponential scatterer (top) and its subset taken from the monotonic region (below).

$$K(f) = \frac{E_{g_{\mathrm{MAX}}} f}{1.225c}. \tag{34}$$

$K(f)$ can then be readily calculated as the frequency f is known in each case, the maximum voltage $E_{g_{\mathrm{MAX}}}$ can be measured from the data, and the speed of sound c is easily calculated from a knowledge of the recorded environmental conditions at the time of measurement.

For the echo from Paraboloid B, the recorded temperature was 22.9 °C, the humidity 36.85%, and pressure 1037.5 mB, resulting in a speed of sound $c=345.0$ m/s. In the 40 kHz case, $E_{g_{\mathrm{MAX}}}$ for the paraboloid's echo was 0.017 78 V, thus the value of $K(f)$ at this frequency is calculated as

$$K(40\ \mathrm{kHz}) = \frac{(0.017\,78)(40\,000)}{1.225(345.0)} = 1.68. \tag{35}$$

For the 60 kHz case, the maximum was found to be 0.02009 V and therefore the value of $K(f)$ is given by

$$K(60\ \mathrm{kHz}) = \frac{(0.020\,09)(60\,000)}{1.225(345.0)} = 2.85. \tag{36}$$

Although it would be preferable to calibrate the model before each new scatterer is tested, it was decided that this would be unnecessary because the impact upon the value of $K(f)$ from the small environmental changes that occur over the entire run of measurements is minimal, given the rather short time span, 10–15 mins, during which the measurements are conducted.

With the calibration constants $K(40\ \mathrm{kHz})$ and $K(60\ \mathrm{kHz})$ now determined, all that remains is for an appropriate time index t to be selected before the matrix elements can be calculated.

B. Selecting an appropriate time index *t*

The ideal choice for time t is within the central region of the echo that lies at least two to three wavelengths into the wave train, bypassing the initial "ramp up" period. In this region the wave is essentially monotonic, monotonicity being a central assumption of Freedman's model and by extension the inverse model. Because the echo is generally sinusoidal in this region, it is reasonable to assume that it can be approximated by a sinusoidal function, the process used to fit this function to the data is discussed next.

Using MATHEMATICA, a subset of the echo data was created that contains only the central monotonic region. The echo data and the corresponding subset, for the exponentially shaped scatterer, are shown in Fig. 6. The structure of the data then needs to be transformed to one suitable for MATHEMATICA's equation fitting function(s). The data are stored in a one-dimensional array, a series that contains the instantaneous amplitude of the echo at discrete time intervals along the full length of the wave train. MATHEMATICA's equation fitting function *NonLinearFit* requires a two-dimensional array as one of its arguments. To create this array, successive elements of the one-dimensional array and their associated time indexes are used as the co-ordinates for successive points within the two-dimensional array. The general form of the equation to be fit is then declared, which in this case is

$$E(t) = A\ \sin(\omega t + \phi) + C. \tag{37}$$

Parameters such as A representing wave amplitude, C representing bias, and ϕ the phase are all determined by *NonLinearFit*. Frequency is a known quantity in each instance. Once these parameters are determined by *NonLinearFit*, a plot of Eq. (37) along with a plot of the error in the fit [the difference between Eq. (37) and the data at each point] is superimposed over a plot of the subset data to ensure that it is a valid fit. An example of such a superposition, for the exponentially shaped body, is shown in Fig. 7. In all cases MATHEMATICA was able to provide an equation with a high correlation to the data.

Next, the condition number of the inverse model's matrix was used to determine the optimal time index t. The condition number of a matrix with respect to inversion is a

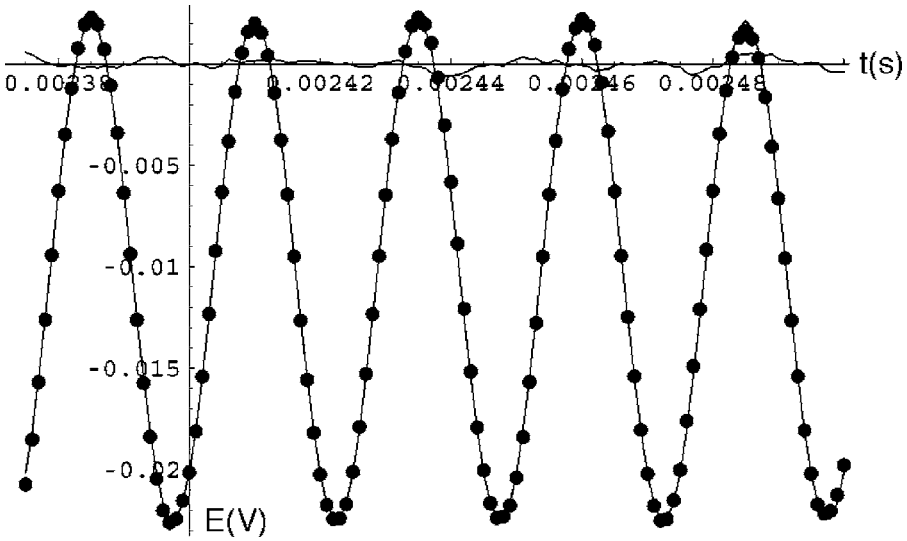


FIG. 7. Superposition of Eq. (37) over the sampled data (solid points) from the exponential scatterer to verify correlation.

measure of how sensitive the matrix is to perturbations. The lower the condition number, the more stable, or well-conditioned, the matrix. Conversely, matrices with large condition numbers are regarded as ill conditioned. In the context of the inverse model, the order of the condition number provides a rough estimate of how many digits of precision are lost from the estimates of $D(n)$ as a result of the matrix inversion. For example, a condition number of 3×10^2, having an order of two, indicates that two digits of precision are lost from the estimates of $D(n)$. So, if the data used only had a precision of three significant figures, after subtracting two digits of precision one is left with an estimate of $D(n)$ precise to only one significant figure, which would be a poor estimate. Clearly, to avoid an outcome with low precision estimates of $D(n)$, either a low-order condition number or data of high precision is required, preferably both. Although sufficient data precision will provide the necessary stability to the model, systematic error in the model will be evident unless the data are also *accurate* to the same resolution. For this reason, the requirements of data accuracy rather than just precision are referenced from this point on. The condition number of the matrix is determined by the MATHEMATICA function *LUDecomposition*.

To determine the time indices at which the condition number is minimized, the condition number of the matrix as a function of time is plotted and it can be seen that there are minima at several points along the wave train, approximately coincident upon echo minima and maxima. See Fig. 8 for an example of this for the exponentially shaped scatterer. The condition number at all minima never falls far below 6×10^2, indicating that the matrix is never well conditioned and will lead to a loss in precision of two digits from any estimate of $D(n)$. This is significant as it severely limits the accuracy of the model, with the major implication of this to be discussed later.

Given the lack of variation in condition number across the various minima, one could apparently select any of the minima in Fig. 8 as the chosen time index, as each will produce a matrix with minimized instability from which a prediction of $D(1)$ can be attempted. However, the signs of the predicted $D(1)$'*s* flip flop between positive and negative with each successive choice of minima. Given that $D(1)$ can *never* be positive, since it is a function of the cross-sectional area in air (i.e., zero) minus the cross-sectional area of the scattering body (which can only be zero or positive) it follows that only the negative valued $D(1)$'*s* can represent real-world solutions. Thus, the predictions for $D(1)$ at each minima are collated, the positive valued solutions rejected and the remaining negative solutions, usually three, averaged with the resulting value used as the model prediction of $D(1)$. The results for all objects are tabulated in Table II and plotted in Fig. 9.

An examination of the results in Table II show a low correlation between the measured and predicted values of $D(1)$. Nevertheless, with the exception of the spheroid and fourth-order bodies, the general trend of the predicted values mirrors that of the calculated ones, despite the obvious differences in magnitude. The measured value for $D(1)$ of the fourth-order scatterer, being negative infinity, could not be shown in Fig. 9. Of particular note is the discrepancy between the measured and predicted values of $D(1)$ for Paraboloid B. Superficially, it would be expected that as the model was calibrated, by virtue of the calculation of $K(f)$, using the data obtained for this object that the measured and predicted values of $D(1)$ should at least match in this instance. However, for the calculation of $K(f)$, the maximum amplitude of the entire received echo was determined and utilized. The prediction of $D(1)$ on the other hand relied upon sampling the wave at several key points and averaging the result. Thus despite coming from the same data, different segments of the data were used for each calculation, thereby accounting for why Paraboloid B was no more likely to have a correlation between theory and measurement than any other object.

C. Error analysis

In the environment under which these measurements were made, short term fluctuations in air pressure from drafts, power supply fans, people walking, etc., could cause variations in the amplitude of the recorded echo by up to

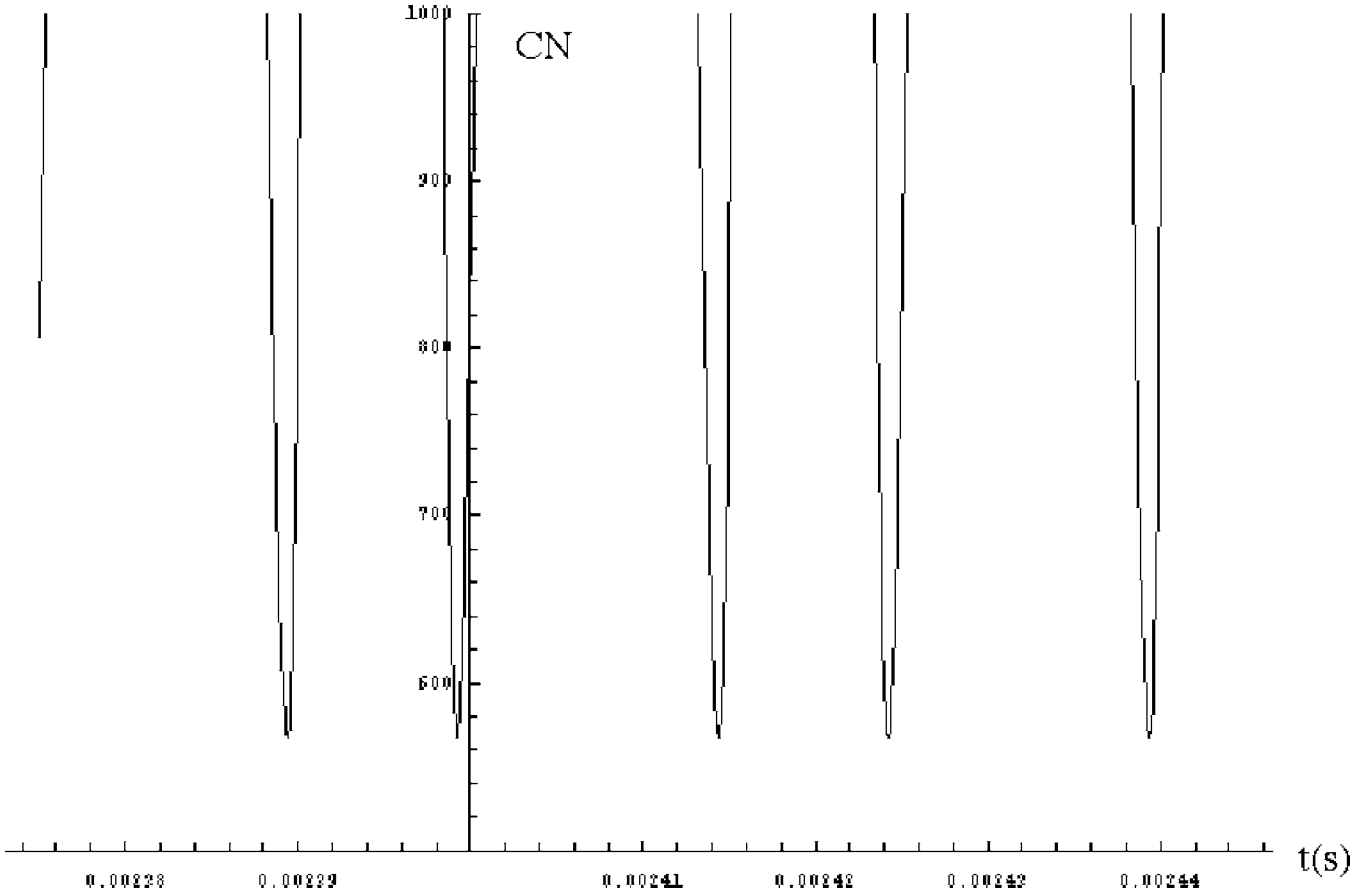


FIG. 8. Condition number of the inverse matrix as a function of time for the exponential scatterer.

10% between any two successive measurements. Specifically, it was found that the standard deviation for any sample of amplitude measurements was approximately 3%. This corresponds to a 68% probability that any random amplitude measurement will be within 3% (one standard deviation) of the population mean and a 95% probability that it is within 6% (two standard deviations) of this mean, the *population mean* here representing the amplitude one would expect in the absence of random noise sources.

Since relatively accurate measures of amplitude were needed for the model, it was advantageous to obtain an estimate of the *population mean of the amplitude*, as this figure represents the amplitude devoid of the influence of random noise. To obtain this, a sample of amplitude measurements was taken, the mean calculated, and the process repeated several times, producing several estimates of the population mean in the process. These estimates constituted a *distribution of means* with a standard deviation that varied depending on the sample size used, with larger sample sizes obviously producing smaller standard deviations and thus a more accurate estimate of the population mean.

TABLE II. Comparison of Measured with predicted values for $D(1)$.

Object	Measured $D(1)$	Predicted $D(1)$
Paraboloid A	$-2.4\pi(-7.54)$	−29.9
Exponential	$-4\pi(-12.6)$	−11.0
Paraboloid B	$-4.9\pi(-15.4)$	−49.2
Spheroid-Paraboloid	$-20\pi(-62.8)$	−228
Ellipsoid	$-40\pi(-126)$	−183
Fourth order	$-\infty$	−54.3

Using a sample size of 50 (i.e., averaging the amplitudes of the echoes from 50 readings) the standard deviation of the distribution of means was found to be only 0.4%. This implies that when 50 echoes are averaged, the estimate of the *population mean of the amplitude* is accurate to within 0.4% in 68% of cases and accurate to within 0.8% in more than 95% of cases. With the error on the amplitude measurements thus being established as just under 1%, the analog-to-digital converter used to sample the echo could be eliminated as a significant source of error since it has a resolution of 12 bits and thus an error of only ±0.02%. Consequently, data can be reported to a maximum of only three significant figures (three digits of precision).

The condition number of just under 6×10^2, having an order of two, results in a loss of precision of two digits from any estimates of $D(n)$. The net effect of these two factors is that predictions of $D(1)$ have a precision and accuracy of only one significant figure, although it needs to be appreciated that the condition number is only a *rough* estimate of the loss of precision. This alone is enough to account for the observed discrepancy between theory and measurement. The possibility of experimental error accounting for the discrepancies was also investigated[4] however. Specifically, the prospect that small variations in the environmental conditions

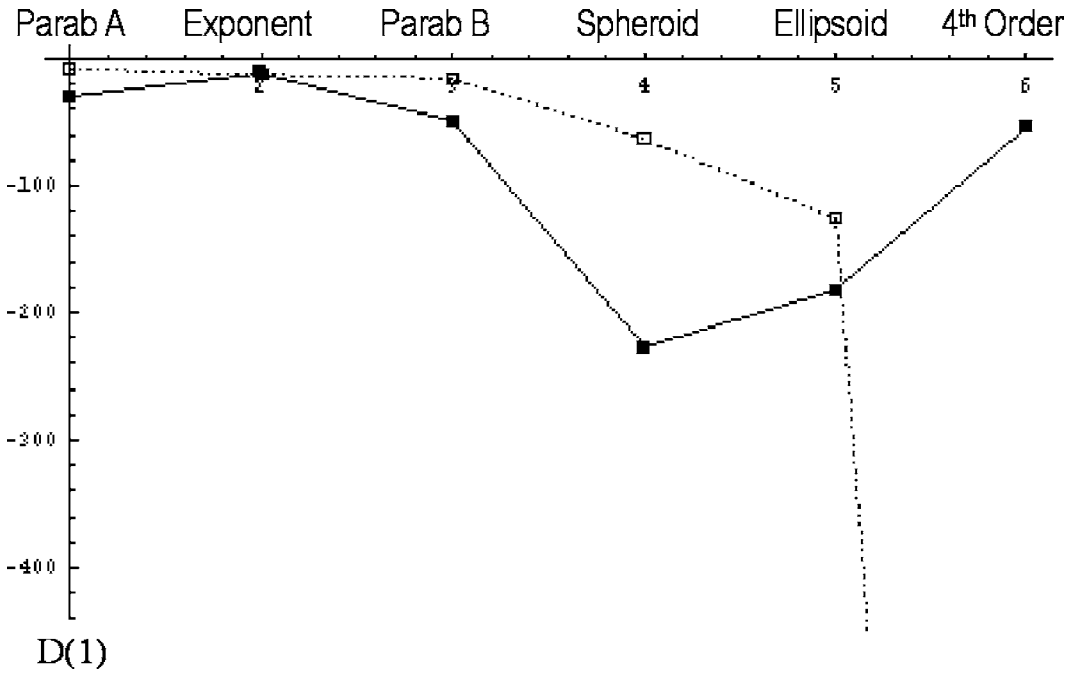


FIG. 9. Comparison of measured (broken line) with predicted (solid line) values for $D(1)$.

could produce significant changes in the predictions of the inverse model. However, even the worst case scenarios for environmental fluctuations could not account for the discrepancies observed.

Last, even discounting experimental error as an explanation for the observed discrepancies, there is little doubt that there are flaws in the inverse model that make accurate prediction difficult. These derive from limitations in the forward model, the same limitations that made it difficult to obtain reasonable predictions for the echoes from cones in the earlier work. In that instance it was not due to an inability to satisfy the assumptions of the forward model, listed earlier, that was the culprit, but rather the failure of the model to account for creeping waves. However, without a detailed analysis of the theory of creeping waves and its integration into Freedman's model, which is beyond the scope of this paper, this impact of this missing element of the model remains conjecture.

However, what is not conjecture is the impact of the ill-conditioned matrix (coupled with insufficient data accuracy) upon the predictions of $D(1)$. The stability of the matrix can only be increased if the condition number is reduced by at least an order of magnitude and the only way to do this would be to reduce the wave number by a similar degree. Transferring the model to a medium like water, for which the speed of sound is more than four times that in air, as well as lowering the frequency of the pulses down to the subultrasonic range (i.e., under 20 kHz) would be enough to reduce the condition number by an order of magnitude. A medium such as steel (for nondestructive testing) with a speed of sound about 17 times that in air, would reduce the condition number by an even greater degree. However, the model would have to be further modified to compensate for the various idiosyncrasies of these mediums and in any event, such medium changes would defeat the purpose of the research, which was to construct the model as an aid for autonomous mobile robotic sensing.

If the model is restricted for use in an air environment only, the condition number cannot be altered to any appreciable degree. Consequently, the only way to counteract the impact of the ill-conditioned matrix would be to use data of exceptional accuracy. Given that the data, with its three significant figures of accuracy, is insufficient for the task, the degree of accuracy required in order for the model to be workable needs to be determined. Ideally, the predictions of $D(1)$ should be accurate to within 1%, implying a measurement accuracy of three significant figures. Given that the large condition number will cause the loss of two significant figures from the prediction of $D(1)$ irrespective of how accurate the data, the data would need to be accurate to at least *five* significant figures to result in a final prediction for $D(1)$ accurate to three figures. An accuracy of this degree can only be achieved if errors in the data are limited to less than 1/10 000.

To do this, the accuracy of the estimate of the population mean of the measured echo amplitudes would have to be increased using a higher sample size to reduce the standard deviation of the distribution of sample means to within 1/10 000. The equation[5] governing this is provided in Eq. (38),

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \tag{38}$$

where $\sigma_{\bar{x}}$ is the standard deviation of the distribution of sample means, σ is the standard deviation of the population (the distribution of measured echo amplitudes), and n is the sample size (i.e., how many echoes are averaged to obtain the estimate).

In practice, reducing the standard deviation of the distribution of sample means to 1/10 000 would still be insufficient, since this would ensure that the estimate of the population mean of the amplitude was accurate to within 1/10 000 in only 68% of cases. The standard deviation would actually need to be reduced to 1/20 000 before the amplitude estimate could safely be stated as accurate to within 1/10 000 in 95% of cases (i.e., two standard deviations), which is acceptable. With measurements of such accuracy, a corresponding increase in analog-to-digital converter (ADC) resolution would be required. The 12 bit ADC currently employed would be insufficient when trying to measure amplitudes to an accuracy of 1/20 000, being only capable of resolving to 1/4600, thus an ADC of at least 15 bits would be required, providing a resolution of 1/32 768.

Using Eq. (38) with the standard deviation of the population of 3%, determined during the earlier work verifying the forward model, and a standard deviation of the distribution of sample means set to a target of 1/20 000, the required sample size is calculated to be 36 000. Thus, almost 36 000 echo samples would need to be averaged before the necessary accuracy could be obtained to satisfy the matrix stability requirements of the inverse model. Given that 50 samples take approximately 30 s, 36 000 samples would take six hours. Clearly this would not be practical. If the accuracy standard of the model is lowered, so that the inverse model is only required to make predictions for $D(1)$ accurate to within 10% (two significant figures), it would still be necessary to average over 3600 echo samples, which would take 36 min. It needs to be pointed out again that because the condition number is only a *rough* estimate of the loss of precision, the calculation of the number of required samples is at best only a ball park figure.

In practice, to sample the echoes for this length of time, whether it is 36 min, or even just 15 min, is not feasible for a couple of reasons. First, as discussed earlier, the Mylar film within the transducer loses its tension if the bias is applied for too long a period. This effect has been measured over several hours but it begins to manifest itself within just a few minutes of continuous use. This problem is currently solved by using the transducer for no more than a minute at a time and then removing the bias for about 60 s. To use the transducer for 36 min continuously (let alone six hours) would inevitably cause the generated echo to reduce in amplitude over the course of the sampling period, as the tension in the film reduces, which would produce a skewed result in the mean. Another reason this is not practical is because environmental changes in temperature, pressure, and humidity over the course of several minutes are non-negligible. Their impact on both time of flight, via their effect on the speed of sound, as well as the absorption coefficient would skew the mean also. It could be argued that the drift in the absorption coefficient can be accounted for by modeling it via calculation of the absorption of sound. However, even with accurate measurements of the environmental conditions, the calculation would only be accurate to within 10%.[6]

These are not insurmountable issues though. With more rigorous laboratory conditions, far stricter environmental controls, as well as a different make or model of transducer, samples could be averaged for a full 36 min without any drift in the mean. Also, if one is willing to transfer the model to either water, for sonar sensing, or steel, for nondestructive testing, where the condition number does not place such high demands on the required accuracy of the data, there may be an application for this model. However, such prospects will have to be left to future workers.

VI. CONCLUSION

In summary, using the extended form of Freedman's forward model, an inverse model was developed that can determine the physical characteristics of an object from a measurement of its echo after insonification and an algorithm proposed for reconstructing the geometry of the scatterer using the predictions of the model. The model was tested with six geometrically simple scatterers at a range of 40 cm using 40 and 60 kHz chirps. A low correlation between the predicted discontinuities and their measured values was obtained, with the primary source of error being determined as the coupling of an ill-conditioned matrix with data of insufficient accuracy. Conditions under which the model should generate accurate results were proposed.

To conclude, this research has raised the possibility for further research into three areas. First, even though the development of the inverse model is mathematically sound, it is based on a forward model that is, to an extent, limited by its failure to account for creeping waves. Further research into the extended forward model from a more theoretical approach could help shed some light onto what could be done to improve the model, the inclusion of creeping wave theory into the model being a starting point. Second, the limitations imposed upon the accuracy of the model in air, by virtue of the ill-conditioned matrix, could be circumvented by modifying the model to work in a different medium, water for sonar-sensing or steel for nondestructive testing. Finally, laboratory environments with greater control of environmental conditions and an ADC with higher resolution could be employed by future workers to verify the model's validity in air.

[1]P. J. McKerrow and N. L. Harper, "Plant acoustic density profile model of CTFM ultrasonic sensing," IEEE Sens. J. **1** (4), 245–255 (2001).
[2]A. Freedman, The formation of Acoustic Echoes in Fluids Doctoral thesis, University of London, 1961, pp. 5–8, 34–58, 95–108.
[3]J. Tsakiris and P. J. McKerrow, "An Analysis of Freedman's 'image pulse' model in air," J. Acoust. Soc. Am. **108** (4), 1602–1613 (2000).
[4]J. Tsakiris, An Inverse Model of Ultrasonic Echolocation Doctoral thesis, University of Wollongong, 2005, pp. 8.24–28.
[5]J. A. Twaite and J. A. Monroe, *Introductory Statistics* (Scott, Foresman and Company, 1979). p. 288.
[6]American National Standard: Method for the Calculation of the Absorption of Sound by the Atmosphere (American Institute of Physics, New York, 1978).

Demonstration at sea of the decomposition-of-the-time-reversal-operator technique[a)]

Charles F. Gaumond,[b)] David M. Fromm, Joseph F. Lingevitch, Richard Menis, Geoffrey F. Edelmann, David C. Calvo, and Elisabeth Kim
Acoustics Division, Naval Research Laboratory, 4555 Overlook Avenue SW, Washington DC 20375-5320

(Received 5 May 2005; revised 14 November 2005; accepted 15 November 2005)

This paper presents a derivation of the time reversal operator decomposition (DORT) using the sonar equation. DORT is inherently a frequency-domain technique, but the derivation is shown in the time-frequency domain to preserve range resolution. The magnitude of the singular values is related to sonar equation parameters. The time spreading of the time-domain back-propagation image is also related to the sonar equation. Noise-free, noise-only, and signal-plus-noise data are considered theoretically. Contamination of the echo singular component by noise is shown quantitatively to be very small at a signal-to-noise ratio of 0 dB. Results are shown from the TREX-04 experiment during April 22 to May 4, 2004 in 94 m deep, shallow water southwest of the Hudson Canyon. Rapid transmission of short, 500 Hz wide linear frequency modulated beams with center frequencies of 750, 1250, 1750, 2250, 2750, and 3250 Hz are used. Degradation caused by a lack of time invariance is found to be small at 750 Hz and nearly complete at 3250 Hz. A back-propagation image at 750 Hz shows focusing on the echo repeater. These results are discussed with comments about further research.
 [DOI: 10.1121/1.2150152]

PACS number(s): 43.60.Tj, 43.60.Fg, 43.30.Vh, 43.30.Re [EJS] Pages: 976–990

I. INTRODUCTION

Recently, advances have occurred in the use of time reversal in ultrasonic systems, notably the development of time reversal operator decomposition (DORT) (Prada *et al.* 1995). DORT can be used to analyze mathematically the process of iterative time reversal that eventually focuses on the strongest scatterer. More importantly, DORT allows the simultaneous determination of all of the resolved scatterers from an analysis of the scattering matrix, or T matrix.

Different approaches to decomposing the scattering matrix have been published (Prada *et al.* 1995; Mast 1997; Montaldo 2001). This paper addresses the use of DORT as a tool to enhance the detection of a scatterer located in the water column in the presence of noise using a ship mounted source and receiver.

This paper is organized as follows. First a brief description of our sonar signal model is given. Then a theoretical description of DORT signal processing as applied to the sonar system is given. After that, results of the sea test are shown. Finally, a conclusion and summary are given.

II. THEORY

A. Signal model

In (Prada *et al.* 1995) the decomposition of the time-reversal operator (DORT) method was developed in the frequency domain for an array of elements that both transmit and receive. The following derivation follows naturally from the original derivation, but is done using sonar terminology with point scatterers and with explicit sources and receivers that can differ from each other (Gaumond *et al.* 2004). In TREX-04, the received signal is noise limited, namely, the dominant impediment to reception of the echo was additive noise rather than reverberation from other distributed scatterers in the environment. The derivation is therefore shown for this case.

Consider the procedure of generating a matrix of scattering data. A set of source elements transmit a sequence of signals and each receiver element records the signal resulting from each transmission. Thus the matrix of scattering data has three dimensions: source index, receiver index, and time. The signal model is the sonar equation (Burdic 1991) in the time domain—where $p(j,l,t)$, $1<j<J$, and $1<l<L$, is the pressure that is received at the jth receiver at position $\vec{r}_j$ in the water column after the lth source transmission, $g(\vec{r}_j,\vec{r}_i,t)$ is the propagation Green's function from position $\vec{r}_i$ to $\vec{r}_j$, $a_m(t)$ is the impulse response of the mth of M point scatterers, and $x_l'(t)$ is the transmitted signal at source position $\vec{r}_l$

$$p(j,l,t) = \sum_{m=1}^{M} g(\vec{r}_j,\vec{r}_m,t) * a(m,t) * g(\vec{r}_m,\vec{r}_l,t) * x'(l,t) + n_l'(\vec{r}_j,t), \quad (1)$$

and the * denotes convolution product in time. The noise signal $n_l'(\vec{r}_j,t)$ has a subscript of the source index to denote the recording of an independent realization of noise at each source transmission. The position of the scatterer can be assumed to lie either in the water column or along the bottom.

[a)]Parts of this material have been presented at the 148th meeting of the Acoustical Society of America [J. Acoust. Soc. Amer.**116**, 2574(A) (2004)].
[b)]Corresponding author; electronic email: charles.gaumond@nrl.navy.mil

B. DORT in the frequency domain

With the assumption that the Green's function and scatterer response are time invariant, a time interval of interest is selected from $p(j,l,t)$ and Fourier transformed into the frequency domain

$$p(j,l,f) = \sum_{m=1}^{M} G(\vec{r}_j,\vec{r}_m,f)A(m,f)G(\vec{r}_m,\vec{r}_l,f)X'(l,f) + N_l'(\vec{r}_j,f) \tag{2}$$

with the convolutions being transformed into multiplications. This element-to-element frequency-domain data matrix can be modified (Lingevitch *et al.* 2002) to account for the use of transmitted and received beams. With $E(l,b,f)$ being the linear transformation between the bth source beam and the lth source element, $F(a,j,t)$ analogously being the linear transformation for the receiver, and $Y(a,b,f)$ being the signal received on beam a from the signal $X(b,f)$ transmitted on beam b

$$Y(a,b,f) = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{m=1}^{M} F(a,j,f)G(\vec{r}_j,\vec{r}_m,f)A(m,f) \times G(\vec{r}_m,\vec{r}_l,f)E(l,b,f)X(b,f) + N_b(a,f). \tag{3}$$

By constraining the analysis to the frequency band of the signal and assuming that the source transmission has been equalized and matched filtered with a unit-magnitude filter to produce a band-limited impulse response with transmitted source level S_0, the signal matrix in Eq. (3) can be written as

$$Y(a,b,f) = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{M=1}^{M} F(a,j,f)G(\vec{r}_j,\vec{r}_m,f)A(m,f) \times G(\vec{r}_m,\vec{r}_l,f)E(l,b,f)S_0 + N_b(a,f). \tag{4}$$

The inverse Fourier transform of $Y(a,b,f)$ is $y(a,b,t)$.

In this formulation, the sonar is not necessarily comprised of a time-reversal array with each element being a source and receiver. The source array might only transmit and the receiver receive. Additionally the number of source elements or beams is not assumed to be the same as the number of receiver elements or beams. Moreover the source and receiver elements are not assumed to be at the same locations. These generalizations render the formulation of the time-reversal matrix $K=Y^HY$ to be symbolic and physically unrealizable because the receivers are not assumed to be able to transmit sound.

Further consideration of Eq. (4) leads to insight on the selection of source and receiver beams, namely, F and E, in the particular operating environment that produces the Green's functions $G(\vec{r}_j,\vec{r}_m,f)$ and $G(\vec{r}_m,\vec{r}_i,f)$ that describe the propagation from the source to the target space $\Omega_T=\{\vec{r}_m\}$. Here, E and F are not assumed to be similarity transformations that preserve the rank and singular-value spectrum of the element-to-element data matrix (Lingevitch *et al.* 2002). Consider the expansion of the target space by the inclusion of well-distributed positions in the volume corresponding to the time interval of interest; the matrix Y is unchanged if these additional points have corresponding target-response elements with values of zero. A propagation model, either numerical or analytical, enables the matrices

$$Q_1(a,\Omega_T) = \sum_{f=f_0}^{f_1}\sum_{m=1}^{M'}\sum_{j=1}^{J} |F(a,j,f)G(j,m,f)|^2 \quad \text{and} \tag{5}$$

$$Q_2(b,\Omega_T) = \sum_{f=f_0}^{f_1}\sum_{m=1}^{M'}\sum_{l=1}^{L} |G(m,l,f)E(l,b,f)|^2, \tag{6}$$

where the frequency summation is over the frequencies in the signal band, to be constructed. Each of these matrices yields a measure of the effectiveness of each beam to project sound into the target space Ω_T.

Similarly the orthogonality of these two beam spaces is seen through the correlation-function evaluated at each frequency

$$C_1(a,a',f) = \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{m=1}^{M'} F(a,j,f)G(j,m,f)G^*(j',m,f) \times F^*(a',j',f) \tag{7}$$

and C_2 defined similarly for the source beams.

Similarly the resolution of each array can be found through simulation by considering the correlation of each sample in Ω_T with each other

$$R_1(m,m',f) = \sum_{j=1}^{J}\sum_{j'=1}^{J}\sum_{a=1}^{A} G^*(j',m',f)F^*(a,j',f)F(a,j,f) \times G(j,m,f) \tag{8}$$

and R_2 defined similarly for the source array. This formula defines resolution in a way consistent with previous references (Prada *et al.* 1995). Note that in general the resolution is a function of two positions as well as frequency.

A qualitative understanding of the functions in Eqs. (5) through (7) is now shown through the use of simple examples. Consider an equally spaced line array in an isotropic, homogeneous, and boundary-free environment with $\Omega_T=\{(r,\cos\theta_m),-1\leqslant\cos\theta_m\leqslant 1\}$, with $G(j,m,f)=(e^{ikr}/r)e^{-ikj\Delta z\cos\theta_m}$, where $k=2\pi f/c$, and $F(a,j,f)=\delta_{aj}$. Clearly each element equally ensonifies each part of Ω_T. The independence of each element can be studied by evaluating Eq. (7)

$$C_1(a,a',f) = \sum_{m=1}^{M'} e^{ik\Delta z(a'-a)\cos\theta_m} \quad \text{which simplifies to} \tag{9}$$

$$C_1(a,a',f) = (e^{i(M'+1)k\Delta z(a'-a)\Delta\cos\theta/2}) \times \frac{\sin[(M'+1)k\Delta z(a'-a)\Delta\cos\theta/2]}{\sin[(k\Delta z(a'-a)\Delta\cos\theta/2)]} \tag{10}$$

with the assumption of $\cos\theta_m=m\Delta\cos\theta$ (Gradstein and Rizikh 1971). This shows that the elements are independent of each other when the condition $k\Delta z(a'-a)\Delta\cos\theta=2\pi$ which constrains the sampling of the target space as well as limiting the choice of frequency to that at which

the array is designed. This makes physical sense because an array with a given length cannot more finely resolve the far field with more closely spaced elements. Adding elements without increasing the length does not increase the resolution of the line array. Likewise for a given length, the array can only resolve a certain equal cosine interval. With the same assumptions, similar results are produced from evaluation of R_1, namely, there are clearly resolved target positions if a set of sampling criteria are met.

The complexity of the former results for a very simple case precludes a presentation of any constructive method for generating suitable beams for a given environment and given target space. Instead, Eqs. (4)–(6) can be used to evaluate beams that are defined in a way to be amenable to modulation, beamforming, transmission, and demodulation.

Next, the general behavior of the signal matrix is treated for three cases: signal dominant, noise dominant, and mixed.

C. Signal dominant case

If the noise is insignificant compared to the echo resulting from the transmitted signal, Eq. (3) can be written in the following matrix form:

$$Y(a,b,f) = \sum_{j=1}^{J}\sum_{l=1}^{L}\sum_{m=1}^{M} F(a,j,f)G(\vec{r}_j,\vec{r}_m,f)A(m,f) \times G(\vec{r}_m,\vec{r}_l,f)E(l,b,f)X(b,f). \quad (11)$$

This equation resembles the singular value decomposition of the data matrix

$$Y(a,b,f) = \sum_{\nu=1}^{M} U(a,\nu,f)\Sigma(\nu,f)V^*(n,\nu,f), \quad (12)$$

where the more common complex transpose has been replaced by a reordering of the indices and a complex conjugation. By definition, the singular vector matrices U and V are orthogonal (Golub and Van Loan 1991). The corresponding matrices in Eq. (11) are also orthogonal if $R_1(m,m',f)$ and $R_2(m,m',f)$ are diagonal over the indices of the actual M scatterers. This means that the actual targets are "resolved" by the set of beams defined by the matrix product of the beam-formed array and the corresponding Green's functions. Namely, the M scatters are resolved by the collection of sources and receivers, respectively, if particular beams $F(a,j,f);1\leqslant a\leqslant M$ and $E(l,b,f);1\leqslant b\leqslant M$ are able to be steered to each of the few individual scatterers. In other words, scatterers are resolved if the source and receiver arrays are able to focus sound on each scatterer independently with a beam. As seen in the simple example in the preceding section, the resolution depends on the frequency, environment, and positions of the sources and receivers.

The similarity between Eqs. (11) and (12) and the orthogonality of $R_1(m,m',f)$ and $R_2(m,m',f)$ implies that there is an identity between the nth individual scatter and one of the singular components, which is also enumerated as the nth here, of the data matrix, namely,

$$\sum_{j=1}^{J}\sum_{l=1}^{L} F(a,j,f)G(\vec{r}_j,\vec{r}_n,f)A(n,f)G(\vec{r}_n,\vec{r}_l,f)E(l,b,f)X(b,f) = U(a,n,f)\Sigma(n,f)V^*(b,n,f), \quad (13)$$

or more simply with the definition of the beam-space Green's functions

$$G_R(a,\vec{r}_n,f) = \sum_{j=1}^{J} F(a,j,f)G(\vec{r}_j,\vec{r}_n,f) \quad \text{and}$$

$$G_S(a,\vec{r}_n,f) = \sum_{l=1}^{L} G(\vec{r}_n,\vec{r}_l,f)E(l,b,f)$$

$$G_R(a,\vec{r}_n,f)A(n,f)G_S(\vec{r}_n,b,f) = U(a,n,f)\Sigma(n,f)V^*(b,n,f). \quad (14)$$

A single component from the singular value decomposition can be used to generate a back-propagation image I_B in the frequency domain by multiplying on the left and right by the simulated, propagating Green's function [Prada *et al.* 1996; Carin *et al.* 2004]. This resulting image is a function of space and frequency:

$$I_{B,n}(\vec{r}_m,f) = \sum_{a=1}^{A}\sum_{b=1}^{B} G_R^*(a,\vec{r}_m,f)U(a,n,f)\Sigma(n,f)V^*(b,n,f) \times G_S^*(\vec{r}_m,b,f), \quad (15)$$

and is evaluated over Ω_T and not just the target positions which are generally unknown. Using the assumed correspondence between the nth singular index and the nth scatterer, Eq. (15) becomes

$$I_{B,n}(\vec{r}_m,f) = \sum_{a=1}^{A}\sum_{b=1}^{B} G_R^*(a,\vec{r}_m,f)G_R(a,\vec{r}_n,f)A(n,f) \times G_S(\vec{r}_m,b,f)G_S^*(\vec{r}_m,b,f)S_0. \quad (16)$$

After Eq. (16) is inverse Fourier transformed into the time domain, the results can be written as

$$i_B(\vec{r}_m,t) = g_{2R}(\vec{r}_m,\vec{r}_n,t) * a_n(t) * g_{2s}(\vec{r}_n,\vec{r}_m,t)S_0, \quad \text{where} \quad (17)$$

$$g_{2R}(\vec{r}_m,\vec{r}_n,t) = \sum_{a=1}^{A} g_R(a,\vec{r}_m,t) * g_R(a,\vec{r}_n,t), \quad (18)$$

$$g_{2S}(\vec{r}_m,\vec{r}_n,t) = \sum_{a=1}^{A} g_S(\vec{r}_m,b,t) * g_S(\vec{r}_n,b,t), \quad \text{and} \quad (19)$$

where $a_n = a(\vec{r}_n,t)$ is the impulse response of a point scatterer at position $\vec{r}_n$. Note that the back-propagation image i_B is a function of both space and time. This time dependence is completely analogous to the time dependence seen in experiments, e.g., [Roux *et al.* 2000].

Thus the time-domain back-propagation image in Eq. (17) equals the convolution of three terms. When the image is evaluated at the position of the scatterer, then the image equals the convolution of (1) the sum of the autocorrelations

TABLE I. The normalized singular spectra of various-sized matrices composed of statistically independent, Gaussian-distributed elements.

Dimension	Singular values
[4, 17]	[0.43, 0.28, 0.18, 0.11]
[5, 17]	[0.38, 0.26, 0.18, 0.12, 0.07]
[6, 17]	[0.34, 0.24, 0.17, 0.12, 0.08, 0.05]

of the propagation from the receiver to the target position [namely, Eq. (18) with $m=n$], (2) the target response function, and (3) the sum of the autocorrelations of the propagation from the source to the target position [namely, Eq. (19) with $m=n$]. Because the propagation autocorrelations are strictly real, the resulting time functions are centered at and symmetric about the origin. An interesting point is that any deviation from temporal symmetry is due to the temporal response of the target $a_n(t)$ that is not convolved with a replica of its response. Thus if the propagation model were accurate enough and if there were target responses of interest, these responses would appear in the time behavior of $i_B(\vec{r}_n,t)$.

The ability to form the back-propagation image depends on the availability of a sufficiently accurate propagating Green's function. If the environment is not known, then the back-propagation image in Eq. (16) cannot be made. However a very limited form of back propagation can be performed using only the $U(a,n,f)$ and $V(b,n,f)$ vectors. Multiplication by these vectors can be interpreted as a back propagation to the position of the scatterer using the following argument. Note that if $G_R(a,r_n,f)$ is evaluated at the position of the nth scatterer, then the propagation vector $G_R(a,r_n,f)$ and corresponding singular vector $U(a,n,f)$ are parallel so that

$$U(a,n,f) = G_R(a,n,f)/\|G_R(a,n,f)\|, \quad 1 < a < A, \tag{20}$$

—and likewise with V—to within a phase shift that depends on the average phases of the two propagator vectors, the phase of the target response and the singular value decomposition (SVD) algorithm. Therefore using Eqs. (15), (16), and (20)

$$\Sigma(n,f) = |I_B(\vec{r}_n,f)|/(\|G_R(a,\vec{r}_n,f)\|\|G_S(b,\vec{r}_n,f)\|), \quad \text{or} \tag{21}$$

$$\Sigma(n,f) = \frac{\left(\sum_{a=1}^{A} |G_R(a,\vec{r}_n,f)|^2 |A(\vec{r}_n,f)| \sum_{b=1}^{B} |G_S(b,\vec{r}_n,f)|^2 S_0\right)}{(\|G_R(a,\vec{r}_n,f)\|\|G_S(b,\vec{r}_n,f)\|)} \tag{22}$$

that shows the singular value is equal to the product of two transmission losses, the source level and the target response. Equation (22) is inverse Fourier transformed to form the time function $\sigma(\nu,t)$, which is not as simply related to physical quantities as the time-domain back-propagation image $i_B(\vec{r}_k,t)$ but is still related to the convolution of the autocorrelations of the propagation and target response. Because $\Sigma(\nu,f)$ is real, by definition of the SVD, $\sigma(\nu,t)$ is symmetric about the origin of time. The time behavior of $\sigma(\nu,t)$ can be expected to reflect the spreading due to propagation and the temporal response of the target.

D. Noise-dominant case

In the noise-dominant case, obtained when there are no significant echoes from any of the transmissions, the frequency-domain data matrix is modeled as a matrix of random numbers

$$Y(a,b,f) = N_b(a,f) \tag{23}$$

that has a singular value spectrum that depends on the particular statistics and size of the matrix. The eigenvalues and singular values of random matrices have been extensively studied (Mehta 1991). The at-sea data are modeled as matrices with statistically independent, Gaussian-distributed elements and with the number of elements comparable to the measured data matrices. This is not at all physics based and ignores important correlations between elements (Kuperman *et al.* 1997). This model also ignores the noise radiated from the ship directly above the vertical line array. On the other hand the use of Gaussian-distributed, statistically independent samples as matrix elements exhibits some important common characteristics of random matrices. These matrices are simulated numerically using a random number generator, the SVD computed, and results are averaged to a specified accuracy. The resulting mean, normalized, squared singular values are given in Table I. These results are accurate to two significant figures. Note that the mean singular values are not equal; the values decay gradually with increasing singular index. This inequality arises because the singular values from each realization are ordered according to size; thus they are ordered before averaging and this ordering generates a spectrum of singular values.

Of course, the elements of the at-sea data matrix are not statistically independent and Gaussian distributed. The resulting singular values differ from those given in Table I. Because the noise in the sea test is measured in a shallow-water environment and the noise arises from distant sources as well as the ship, deviations from the simple model are expected. However, the spectrum of singular values with a trend of decay is indicative of a random matrix. Also, if there were a component of the noise that originated in a particular direction, then the distribution of singular values would be skewed by the signal energy of that single component, which would be concentrated primarily into a single component. This is shown in the next section.

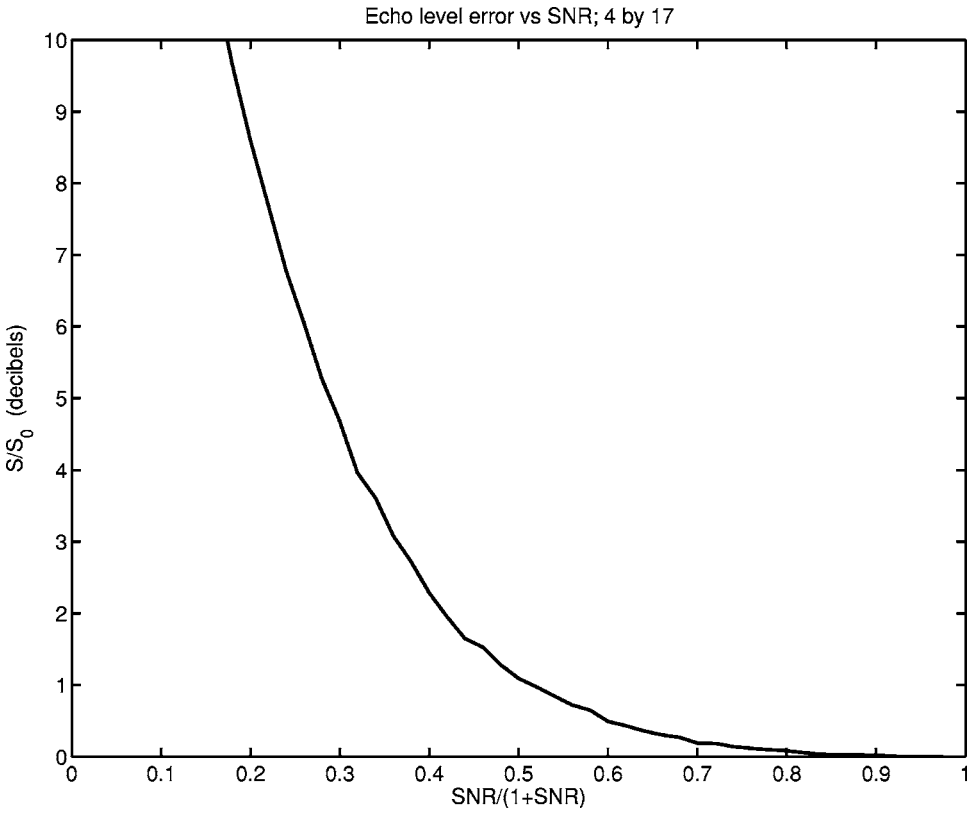


FIG. 1. The error of the signal level is shown as a function of SNR/(1+SNR). The noise matrix has statistically independent, Gaussian-distributed random elements. Averages were found to two significant figures. Note that the error in echo level is less than 1 dB at SNR=1.

E. Signal plus noise case

Consider the next level of complexity that roughly corresponds to the noise-limited data taken in the experiment, namely, a data matrix that is composed of a rank-1 signal matrix and a full rank noise matrix

$$Y(a,b) = U_0(a)S_0V_0^*(b) + N(a,b), \tag{24}$$

where the vectors U_0 and V_0 are normalized vectors, S_0 is the scalar echo-level, and $N(a,b)$ is matrix comprised of uncorrelated, Gaussian-distributed elements, as above in Eq. (23), with a Frobenius norm $\|N(a,b)\|$. This represents the structure of a data matrix with a single scatterer with additive noise. As in the preceding section, the assumption of uncorrelated elements introduces more degrees of freedom than would be found from a numerical simulation of a sonar experiment in shallow-water acoustics, where the signal space is constrained by the acoustics of the particular environment. An estimate of the signal is obtained from the data matrix from the SVD of the data, namely, $Y(a,b) = \Sigma_{\nu=1}^{A} U(a,\nu)\Sigma(\nu)V^*(b,\nu)$, $a<b$, by assigning the first singular value to the estimated echo level $\hat{\Sigma}_0=\Sigma(1)$, receiver steering vector $\hat{U}_0(a)=U(a,1)$ and source steering vector $\hat{V}_0(b)=V(b,1)$. The error in estimating the echo level is defined as $E\{\Sigma(1)/S_0\}$, where $E\{\}$ denotes the expectation value or the average. The accuracy in estimating the receiver and source steering vectors is defined as $E\{\Sigma_{a=1}^{A}U^*(a,1)U_0(a)\Sigma_{b=1}^{B}V(b,1)V_0^*(b)\}$. The product of the receiver and source steering vectors is used because the phase or sign of each vector is defined arbitrarily in the SVD. These errors are found to a specified accuracy by numerical simulation with a sufficient number of realizations. This was done for a variety of values of signal-to-noise-ratio SNR $=S_0/\|N\|$. The results are plotted with respect to the ratio SNR/(1+SNR) so that both asymptotes—full signal and full noise—can be shown easily and compactly.

Figure 1 shows the echo-level error in decibels for the case of four (4) source beams and seventeen (17) beams. Note that for SNR=1, or 0 dB, the error is less than one decibel which is quite low.

Figure 2 shows the steering vector accuracy on a linear scale. This error is approximately 0.9 at an SNR=1. This accuracy has an effect on back-propagated images; an accuracy of 0.9 means that the peak of the image is decreased only by a factor of 0.9. However, 0.1 of the signal is spread among the other singular values. This spreading of a large singular value can be very visible in back-propagation images because this small amount of energy is added to ranges and depths where there was very little or none before.

Figure 3 shows the squared, normalized singular values $\Sigma(\nu)^2/|\Sigma|^2$ as a function of SNR/(1+SNR). This figure shows the change of the singular spectrum from a full rank, random-noise result near SNR/(1+SNR)=0 to a rank one spectrum near SNR/(1+SNR)=1.

F. Time-frequency domain

In the previous sections, a set of data was analyzed in the frequency domain. In this section an explicit time-frequency notation is shown that can be more easily used for sonar applications. This was not done in the beginning to alleviate the use of an additional variable to the already long list of variables. This time-frequency analysis is imple-

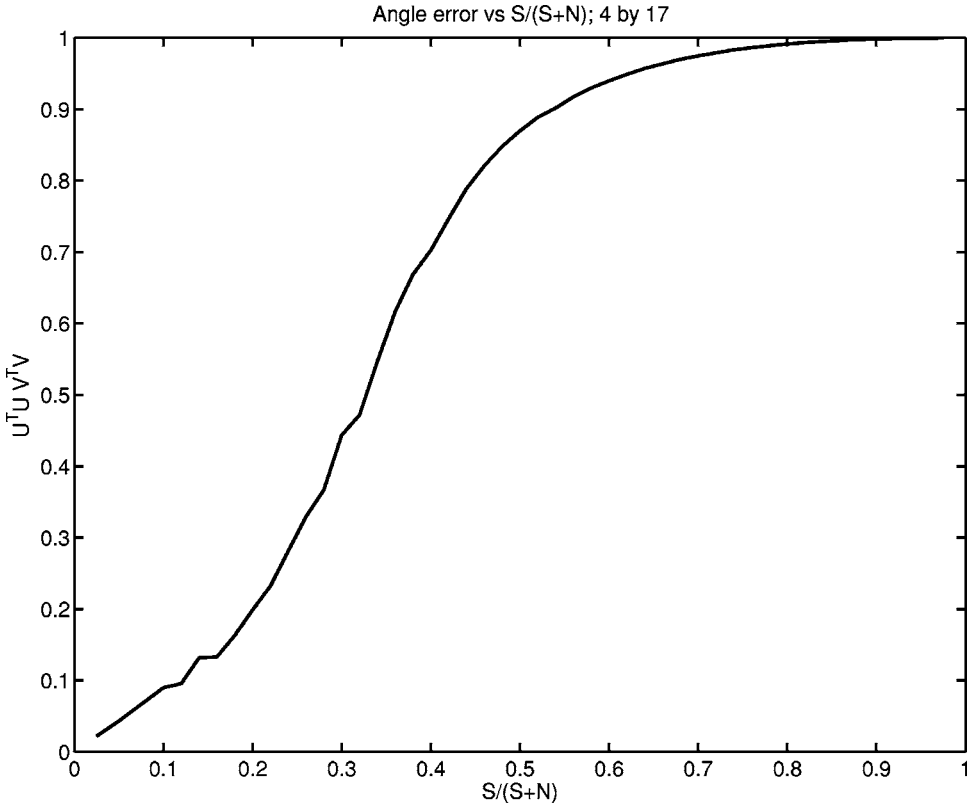


FIG. 2. The accuracy of the steering vectors as a function of SNR/(1+SNR). The product of accuracies of the receiver and source steering angles is only approximately 0.9 at SNR=1.

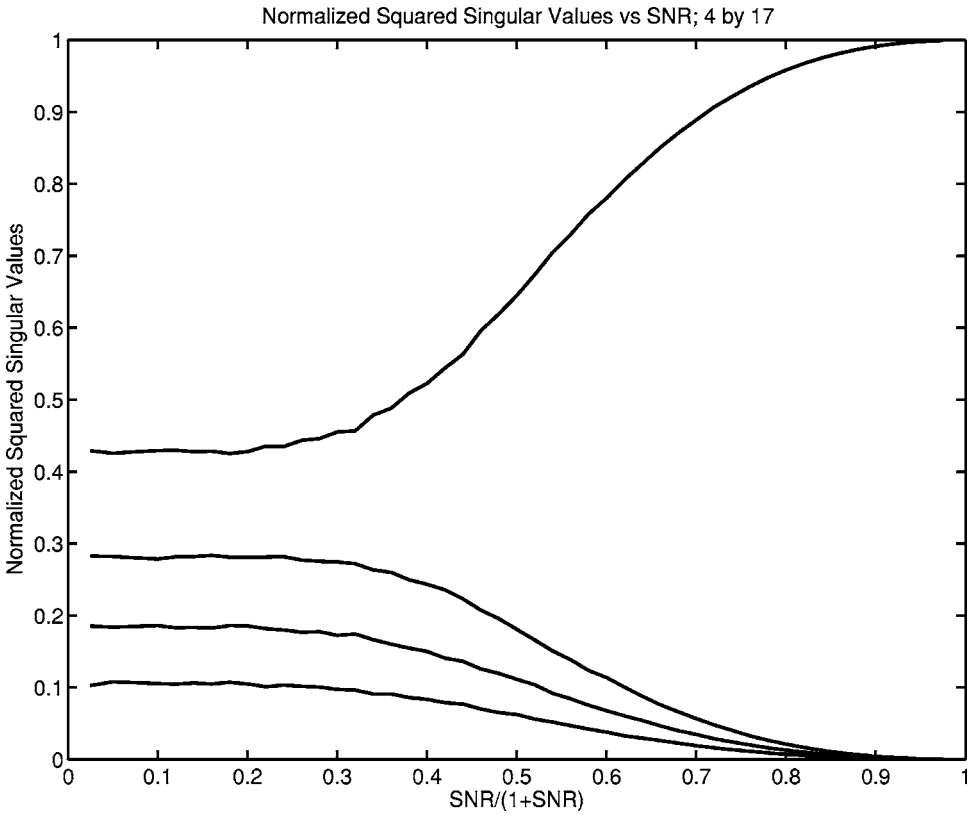


FIG. 3. The values of the singular values as a function of SNR/(1+SNR). The singular values have been normalized. Note that at very low SNR the singular value spectrum is that of a matrix with statistically independent, Gaussian-distributed elements. At very high values of SNR the spectrum has only one nonzero value.

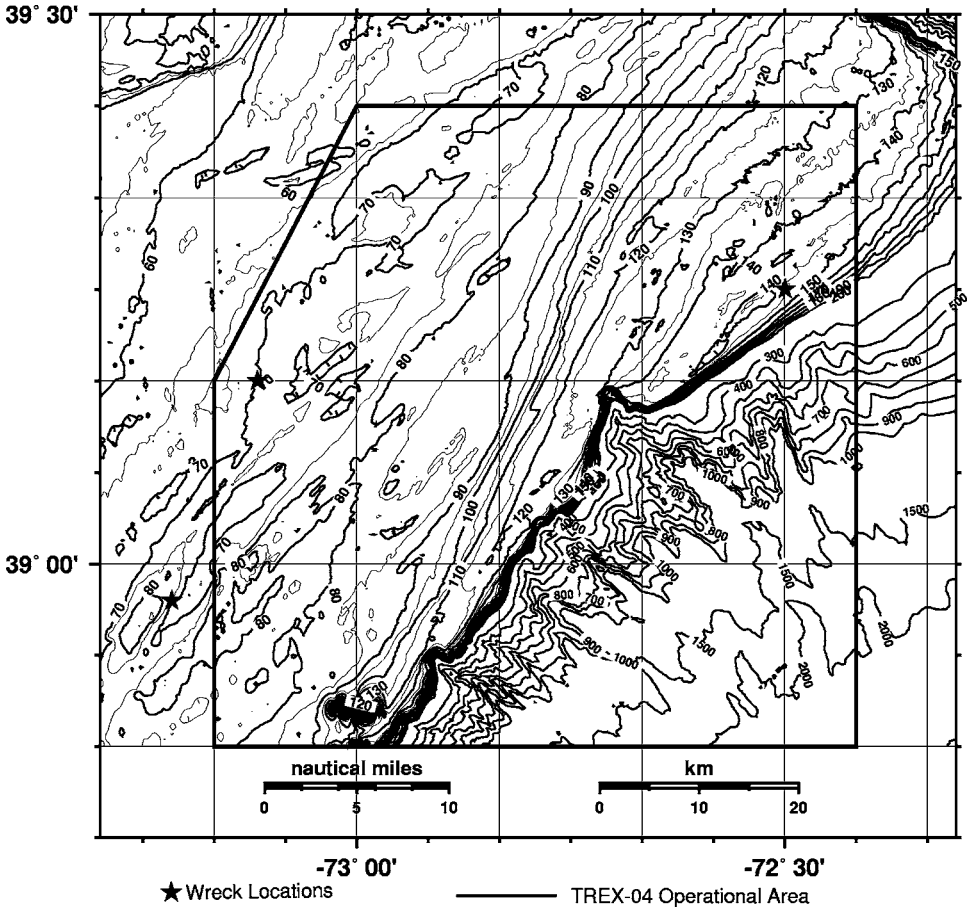


FIG. 4. A map of the experimental area. The experiment was performed in the vicinity of 39.24°N and 72.79°W.

mented using a sequence of windows separating the time samples of the signal matrix into D epochs. The windows obey

$$1 = \sum_{d=1}^{D} w(t - dT), \quad 0 < t < (D-1)T, \tag{25}$$

where T is the time shift associated with the window and dT is the epoch time. Several windows obey this criterion, e.g., 50% overlapping triangular window or nonoverlapping rectangular window.

The application of this decomposition on $y(a,b,t)$, the inverse Fourier transform of $Y(a,b,f)$ in Eq. (4), yields the signal matrix

$$y(a,b,d,t) = y(a,b,t)w(t - dT), \tag{26}$$

where the time variable t extends over the duration of the time-window. The index d denotes the epoch time dT from which range can be estimated. The Fourier transform of $y(a,b,d,t)$ over short-time yields the signal matrix $Y(a,b,d,f)$. The DORT algorithm is then applied to each epoch and frequency independently

$$Y(a,b,d,f) = \sum_{\nu=1}^{\mathrm{MIN}(A,B)} U(a,\nu,d,f)\Sigma(\nu,d,f)V^{*}(b,\nu,d,f). \tag{27}$$

Thus for each epoch and frequency there is a set of singular values $\Sigma(\nu,d,f)$ where ν is the singular index, d the epoch index, and f the frequency index.

Remembering that $\Sigma(\nu,d,f)$ is proportional to the back propagation image focused on the position of the target as shown in Eq. (22), the inverse Fourier transform $\sigma(\nu,d,t)$ is quantitatively related to the impulse response of the target convolved with the autocorrelations of the propagation to the target from the source and from the receiver. This function is expected therefore to be a strongly peaked function and therefore $\sigma_{\mathrm{peak}}(\nu,d) = \mathrm{MAX}[\sigma(\nu,d,t)]$ can be plotted as an A-scan trace for each singular index ν, with each trace related to the portion of the signal matrix due to each resolvable scatterer.

III. EXPERIMENT

A. General description

The time reversal experiment (TREX-04) was conducted southwest of the Hudson Canyon as shown in Fig. 4. The research vessel (R/V) Cape Henlopen was moored at 39.24°N and 72.79°W on the 92 m contour from April 22 to May 4, 2004. From this ship, the Naval Research Laboratory (NRL) 64-element source-receiver array (SRA) was deployed from the A frame on the stern in a vertical configuration. The SRA operates between 500 and 3500 Hz with element spacing of 1.25 m. The R/V Endeavor drifted with an echo-repeater system consisting of either a single Raytheon XF4 or an ITC 200 source. The XF4 was used for frequencies from 500 to 2500 Hz; the ITC 200 covered from 2500 to 3500 Hz. The data shown in this paper was taken on May 1 and May 2. The wind speed varied from 11 to 22 kn.

TABLE II. Sequence of transmissions.

Frequency (Hz)	Chip type	Chip duration (msec)	Chip interval (msec)	Steering angle (+degrees=up)
0.5–1.0	LFM	250	250	[0 0 0 0]
0.5–1.0	LFM	250	250	[5 0 −5 −10]
1.0–1.5	LFM	250	250	[0 0 0 0]
1.5–2.0	LFM	250	250	[0 0 0 0]
2.0–2.5	LFM	250	250	[0 0 0 0]
2.5–3.0	LFM	250	250	[0 0 0 0]
3.0–3.5	LFM	250	250	[0 0 0 0]
3.0–3.5	LFM	250	250	[5 0 −5 −10]

DORT processing requires a set of source transmissions and receiver recordings taken quickly enough that time invariance can be assumed. This stringent criterion may be difficult to obtain in the ocean especially with ship-mounted equipment. Temporal changes in the Green's function cause erosion of the rank-1 structure of scattering from an isolated, single point. Several methods are possible for transmitting a sequence of source signals. The most obvious is transmitting the first source signal, then recording the response from it, and repeating that sequence with the second signal. This transmit and record sequence can take many tens of seconds. Another way to transmit the sequence would be to use closely spaced frequencies for each element or beam [Folegot *et al.* 2004]. Another alternative is the use of code division multiple access (CDMA) which uses an approximately orthogonal wave form for each source element or beam. In this experiment, a time division multiplexing (TDM) scheme was used with a transmitted sequence of linear frequency modulated (LFM) beams lasting a total of one (1) sec. Thus the data matrix is filled within an interval of 1 sec. Table II shows the signals that were transmitted to produce the results shown in this paper. Each 250 msec LFM is immediately followed by the following LFM. There is a 30 sec cycle time due to duty cycle restrictions. Note that a sequence of broadside signals, steering angle=[0° 0° 0° 0°], is transmitted in each frequency band. This sequence of identical signals is transmitted to test the time invariance over the duration of the transmission, which was 1 sec (namely, four times 250 msec).

B. Signal processing

The signals $p_0(m,t-nT)$, $m=1,\ldots,64$, $n=0,1,2,3$, and $T=250$ msec are received on the 64-channel SRA with a sample rate of 10 129 Hz. These signals are then coherently matched filtered with the appropriate LFM and demodulated appropriately. Figure 5 shows the sum of the squared channels versus time at this stage. There are four sharp peaks in the signal shown in Fig. 5. These are echoes from the sequence of broadside source transmissions. The range of the

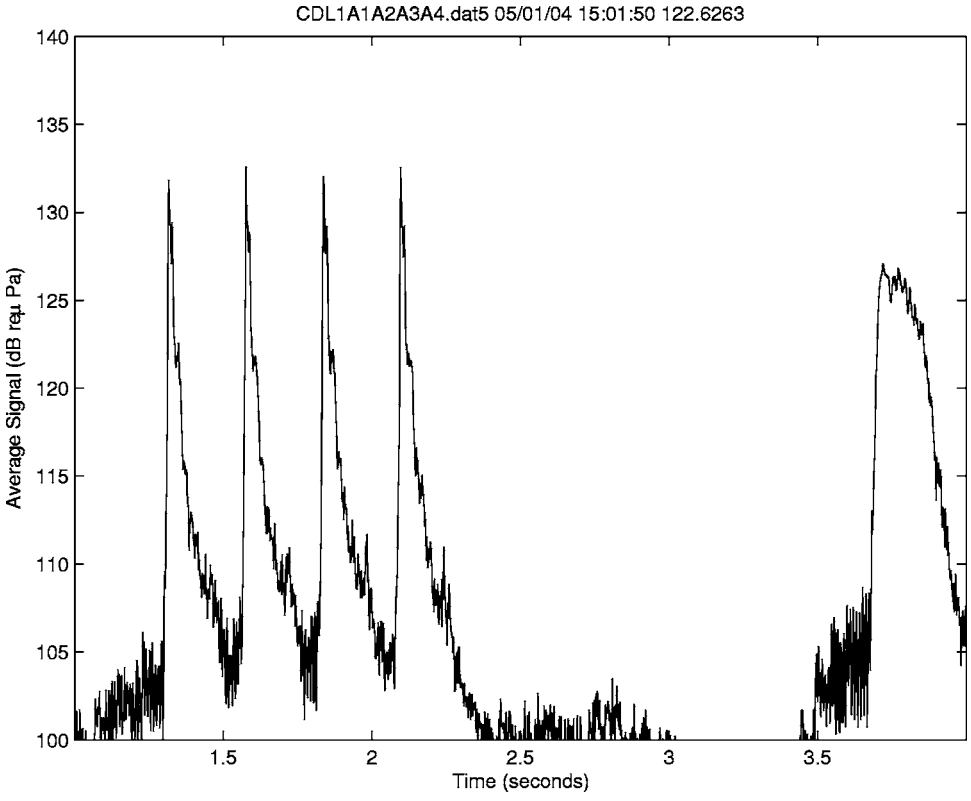


FIG. 5. The signal received on the SRA produced by matched filtering, squaring, and summing the signal from each element. The first four peaks are the responses from a sequence of four broadside transmissions. The signal at the right was transmitted for a different experiment and should be disregarded.

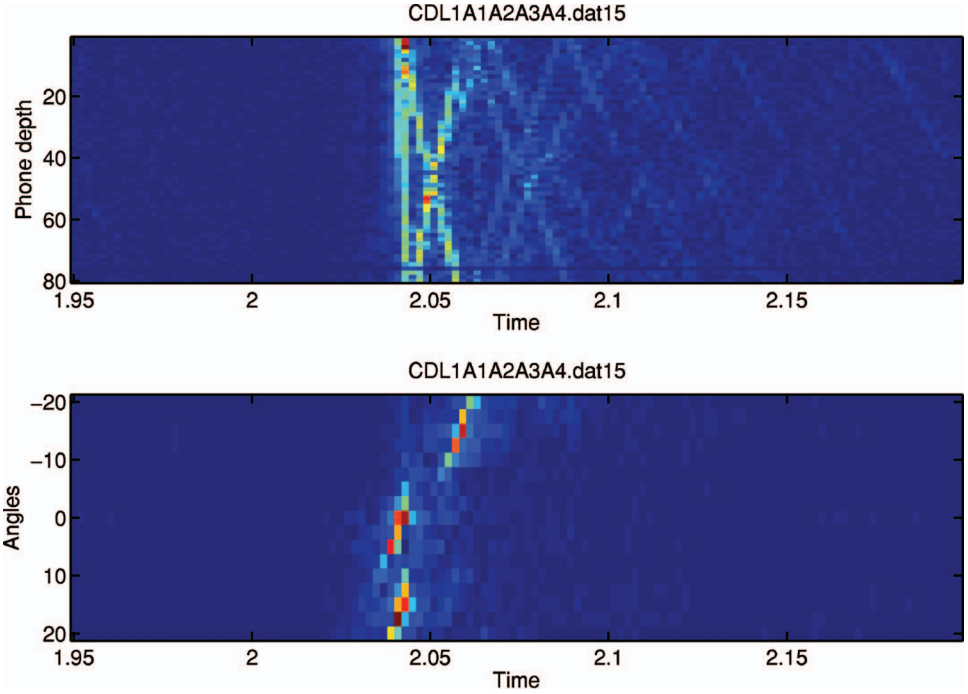


FIG. 6. The matched-filtered signal is displayed as recorded in the upper panel as a function of phone depth in meters and time in seconds. The lower panel shows the signal transformed into angle-time (beam-time) space with the angles in units of degrees and time in seconds.

echo repeater was approximately 0.5 km; the additional time delay of the echo repeater has translated the echo to be later in time. Note that the signal-to-noise ratio is very large and that each signal decays to approximately the noise level within the 250 msec ping separation. Thus there is a negligible contribution from the late-time response of preceding echoes in the second through fourth echo.

In order to decrease the dimensionality of the signal, the 64 channels are beamformed into 17 equally spaced beams, $[-20\ -17.5,\ldots,17.5\ 20]$. Almost all of the signal energy is concentrated into that angular range. An example of one received echo is shown in Fig. 6. The upper panel shows the data as a function of position along the array and time. Note that the coarse sampling in the middle is due to several dead phones. The wave-front structure of the signal is caused by the pointlike echo repeater that is approximately 0.5 km away. These wave fronts are mapped into smeared peaks in the lower panel that shows the signal in the angle-time domain. The first wave front arrives nearly broadside and is mapped into a beam directed near 0°. The weak late-time arrivals are not included; this corresponds to the portion of the signal that is more than 20 dB down from the peak as seen in Fig. 5. The color scale in Fig. 6 is linear.

In order to form the four source channels for each receiver channel in the data matrix the array of signals is time delayed in units of the ping separation, namely,

$$y(a,b,t) = y_0(a,t-bT), \quad b = 0,1,2,3, \tag{28}$$

so that one set of echoes is time aligned. The four source channels are shown in Fig. 7.

After the signals are time aligned, a 250 msec window is placed so that the four responses are contained within it and the signal is Fourier transformed and processed as described in Sec. II.

C. Coherence

Consider the quantification of the time invariance obtained in this experiment. In order to quantify the time invariance, four identical signals are transmitted. With the notation for a time-varying Green's function from the echo-repeater $\vec{r}_E$ to each receiver $\vec{r}_j$ at frequency f and time t_b as $G(\vec{r}_j,\vec{r}_E,f;t_b)$, the data matrix resulting from the fourfold transmission of source beam 1 is

$$Y(a,b,f) = \sum_{j=1}^{J}\sum_{l=1}^{L} F(a,j,f)G(\vec{r}_j,\vec{r}_E,f;t_b)A(E,f) \times G(\vec{r}_E,\vec{r}_l,f;t_b)E(l,1,f)S_0 + N_b(a,f) \tag{29}$$

and the matrix is rank one if the Green's functions are the same at each of the four times. (Note that a change in the Green's functions can arise due to motion of the source, echo repeater, receiver, or the oceanic medium.) The data is then processed using the DORT procedure in Sec. II with a window that completely contains the four time-aligned echoes. However, before proceeding to the discussion of the DORT-processed data, consider first the normalized singular spectrum of ocean noise. A time window with no echo signal is chosen and the noise signal analyzed using the same time shifts and DORT decomposition. The resulting normalized singular values are plotted as functions of center frequency in Fig. 8. The means of each set of singular values are plotted also. Note that the mean normalized spectrum, [0.59 0.22 0.12 0.06], differs from that of a 4×17 Gaussian matrix, namely [0.43 0.28 0.18 0.11], as shown in Table I. This disparity probably arises from the arrival structure of the noise field that comes from distant sources, the R/V Endeavor and the R/V Cape Henlopen that is located over the array. These resulting singular values represent the worst coherence possible in the sys-

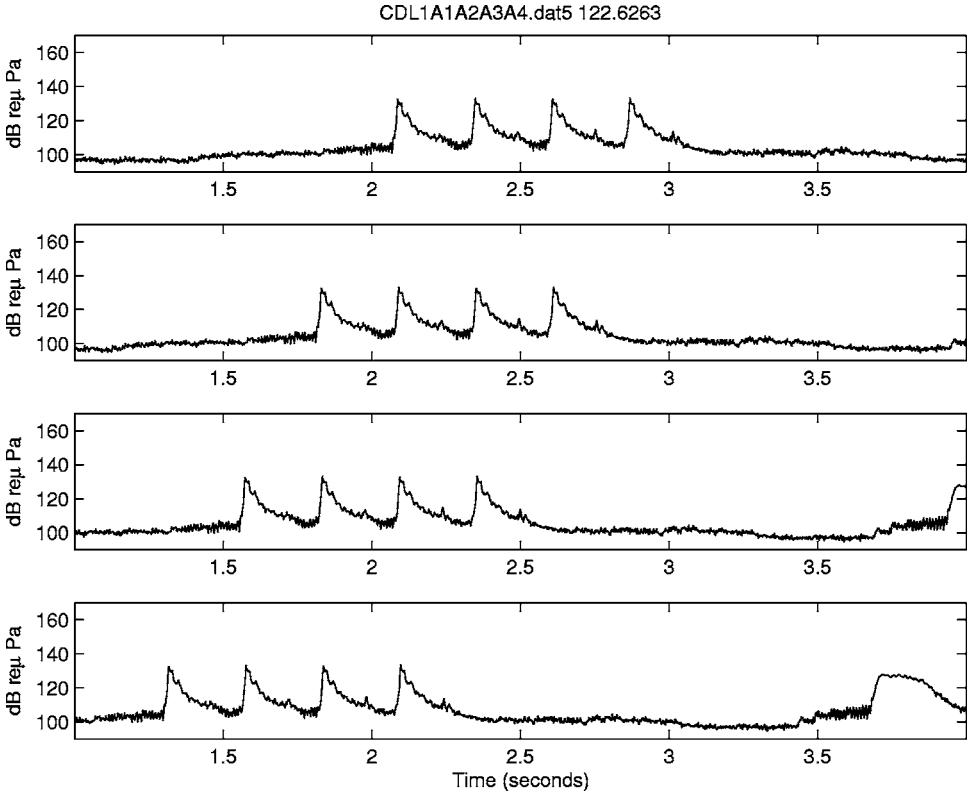


FIG. 7. The result of time shifting the signal by 250 msec so a set of four peaks are aligned. Each panel from top to bottom is given a different source index. Only in the time window from 2.00 to 2.25 s are all four source signals present.

tem. Thus the coherence analysis of the echo-repeater signals is expected to produce a highest normalized singular value between the value of 0.59 found from ocean noise and 1.00, found theoretically for perfect coherence and no noise.

Knowing the expected bounds of performance, DORT processing is performed on one signal from each frequency band at different ranges and wind speeds as the data permit. The largest normalized singular value is then plotted in Fig. 9 versus center frequency, winds speed, and range. This scatter plot shows that there is a clear dependence of coherence on the frequency as opposed to range and wind speed (given

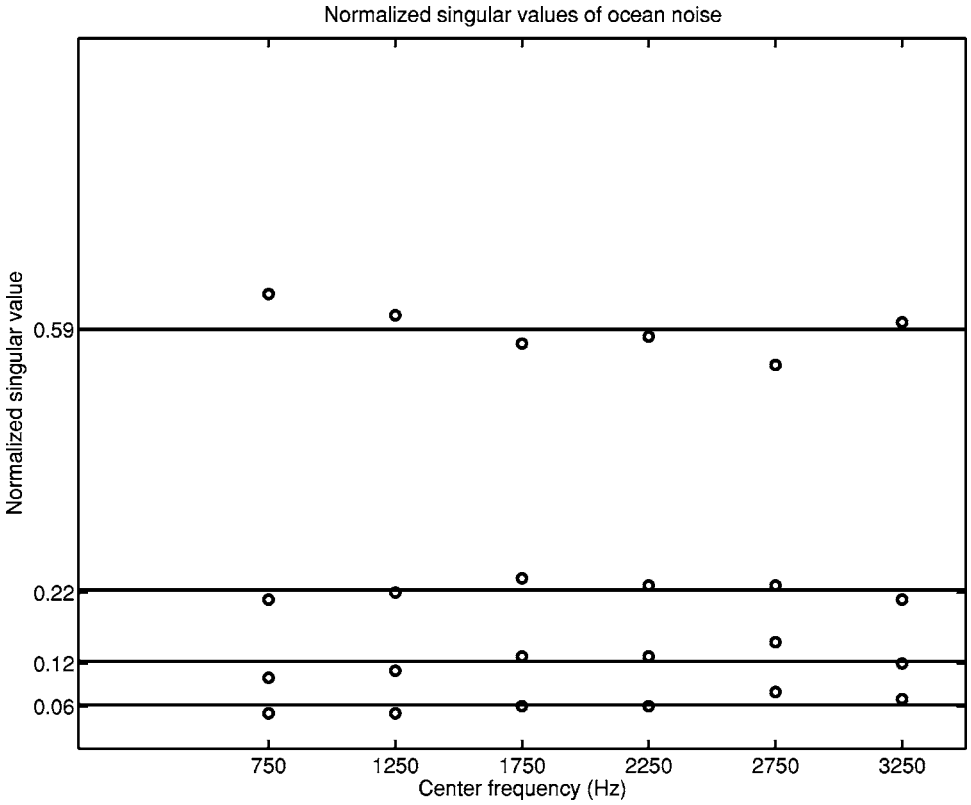


FIG. 8. The four singular values of oceanic noise recorded and processed exactly the same as the echoes. The singular values are normalized. There is little frequency dependence on the singular values. The average over frequency of each singular value is displayed on the plot.

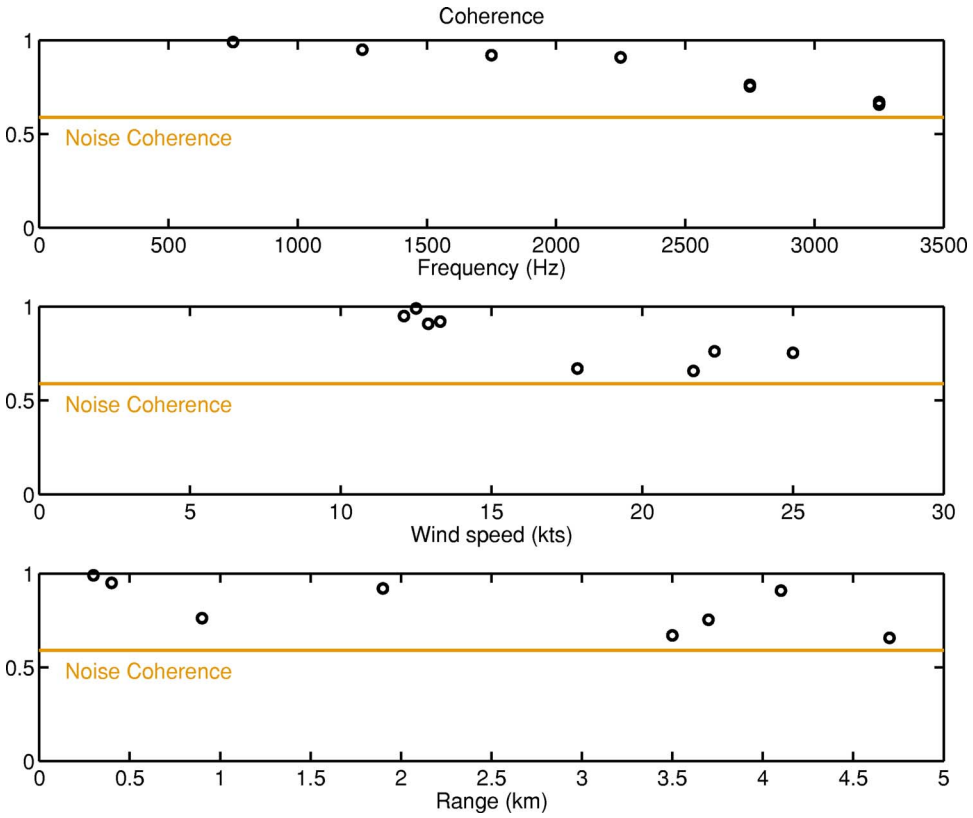


FIG. 9. (Color online) The first, and highest, normalized singular value resulting from four broadside transmissions in the various 500 Hz bands. The first singular value is plotted with respect to frequency in the top plot, with respect to wind speed in the middle and with respect to range of the echo/repeater in the bottom plot.

the narrow range of wind speeds). This dependence is probably due to the heaving of the ship and the vertical motion of the array during transmission and recording of the signal. The motion during the one second interval has a considerable effect at the highest frequency band. At the highest frequency band, the largest singular value approaches that from ocean noise. The value in the lowest frequency band is very close to one, which is the value expected for perfect coherence and very high SNR. This ideal case is shown in Fig. 3 by the uppermost curve for values of SNR/(1+SNR) close to one.

D. DORT analysis of signals

The previous section addressed the potential coherence of the system for use with DORT processing. The quality of the coherence was assessed by the magnitude of the largest normalized singular value with a transmitted signal with a very large signal-to-noise ratio. This is a measure of the isolation of a signal into a single component.

There are two other aspects of each signal component beside the magnitude of the singular value, namely, the receiver and source singular (steering) vectors that are functions of either frequency or time. The quality of these components depends on their pointing to the same spatial region over frequency and to their compact extent in space time. These qualities cannot be investigated without a back-propagation model that shows the propagation of energy from the steered source and receiver arrays. This was done for the lowest frequency band using Eq. (15) and inverse Fourier transforming into the time domain to generate the coherent back-propagation image $i_B(\vec{r}_k,t)$. The required Green's functions were calculated using a version of the range-dependent acoustic model (RAM) program (Collins 1993) where the self-starter was modified to insert the starting field for an entire vertical array of source elements, including spatial shading and phase delays for steering the array. The 64 array elements were placed every 1.25 m starting at 1.25 m from the ocean surface. The spatial shading was uniform and the phase delays were standard for a delay-and-sum plane-wave beamformer. The Green's functions were calculated every 4 Hz from 500 to 1000 Hz for the required source and receiver beam steering directions. Range-dependent bathymetry between the SRA and the echo-repeater locations was extracted from a high-resolution database (Goff *et al.* 1999). For the results presented, the depth varied from 92.1 to 91.6 over the 0.5 km of range. The sound speed profile is given in Table III. It was derived from a combination of expendable bathythermograph (XBT), thermistor, and conductivity, temperature and depth (CTD) data collected throughout the experiment. The geoacoustic data (Gauss 2003) is given in Table IV. Finally, a consistent set of algorithm control parameters were used for all RAM runs; specifically, dz=0.025 m, dr=1.0 m, and the number of Pade coefficients set to 5.

The coherent back-propagation image $i_B(\vec{r}_k,t)$ is seen in Sec. II C to be a function of range, depth, and time. The results are computed for a set of ranges near the range expected from the global satellite positioning (GPS) measurements taken on each ship. The peak response is shown in Fig. 10 for range=0.5 km as a function of depth and time. The back projection for each of the four singular components are shown in the four subplots, with the top having the largest singular value and the lower subplots in descending order. The color scales of each subplot have a 30 dB range from 70 to 100 dB (arbitrary units). The position of the echo

TABLE III. Sound speed profile.

Depth (m)	Sound speed (m/s)
0	1482.7
6.75	1482.7
10.5	1481.4
14.25	1481.5
18	1480.9
21.75	1481.0
25.5	1481.3
29.25	1481.6
33	1481.7
36.75	1481.3
40.5	1483.3
44.25	1483.2
48	1483.5
51.75	1483.7
55.5	1483.6
59.25	1483.5
63	1483.5
66.75	1483.5
70.5	1483.6
74.25	1483.7
78	1483.7
81.75	1483.8
92	1483.8

repeater is clearly visible in the plot at a depth of approximately 30 m and a time delay of 0.08 sec. The depth of the echo repeater was measured to be 40 m with a depth gauge.

In order to compare the spreading in depth and time, an ideal back-propagation image was simulated for an impulsive target location with range=0.5 km and depth=40 m. The ideal data is created using the same Green's functions by first calculating a received time series from the target for each combination of source and receiver beams. These time series are then processed the same as the experimental data. The spatial and temporal impulse responses of the source and receiver arrays in this environment are shown in the four subplots of Fig. 11. As in Fig. 10, the top subplot corresponds to the largest singular values. The range of the color scales is again 30 dB. Because the ideal data is completely noise free, there is no energy present for singular values 2–4. Note that the ideal spatial focus is very sharp with a spread of only a few meters in depth. The time spread is also quite short with a well-defined peak. The deviations of the image in Fig. 10 from ideal are assumed to be due to environmental mismatch because the singular values for this case showed a very high level of coherence.

The time delay shown in Fig. 11 has been time shifted by 0.125 sec. so that the peak lies in the middle of the plot. Because the ideal echoes have exactly the same time delays of the back-propagation Green's function, the time delay of the peak would occur at a time delay of zero. The nonzero time delay of the back-propagated data in Fig. 10 should not be regarded as significant because there is an unknown time delay generated by the echo repeater. The echo repeater receives the incident signal, stores the signal for approximately five seconds, and then retransmits the signal. There is considerable time jitter in the echo repeater. This would not occur for a passive target when the time delay would be indicative of a mismatch between the data and the back-propagation Green's functions.

The data collected generally lacked evidence of bottom reverberation that rises above the ambient noise level. Thus, in Fig. 10 there are no hints of spatially isolated layer bottom reverberation in the back propagation of singular component 2. The lack of reverberation could be due to the modest source level (especially at the lower frequencies), the particular steering angles used for the source transmissions, the warm-water incursions near the bottom that tended to make the sound speed profile upward refracting, or low bottom backscattering strengths. In order to test the efficacy of spatially separating bottom reverberation from this configuration, an ideal data set is generated for targets at the depths of the echo repeater and the bottom. Again the ideal data is constructed from the back-propagation Green's functions. The resulting images are shown in Fig. 12. Note that the separation of these two targets is not complete with a smeared target component being mixed with a strong bottom component in the top subplot. The second subplot contains primarily the target component with a small portion of the bottom component that has leaked into the second singular component. The mixing, or leakage, of the two components in the time domain is caused primarily by frequency-

TABLE IV. Geoacoustic parameters.

Depth (m)	Sound speed (m/s)	Density (g/cc)	Attenuation (db/L)
0	1635	1.5	0.57225
94	1635.96	1.5	0.57225
104	1635.27	1.5	0.57225
114	1635.27	1.5	0.57225
144	1635.27	1.52	0.57225
194	1635.27	1.59	0.57225
244	1690.27	1.66	0.57225
294	1743.27	1.73	0.57225
300	2200	2.15	0.57225
400	2200	2.15	10

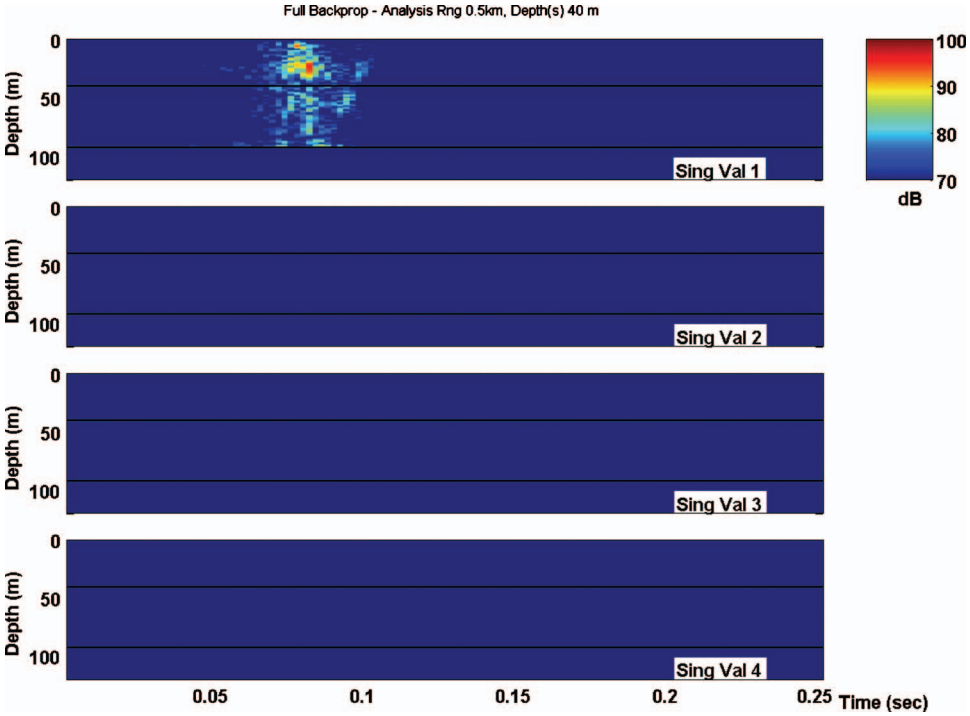


FIG. 10. The back-propagation image $i_B(\vec{r},t)$ of one ping from 500 to 1000 Hz versus depth and time with range=0.5 km. Each panel displays the back-propagation image with an identical color scale with 30 dB dynamic range. The top panel displays the image from the largest singular vectors that are associated with the echo-repeater signal. The expected depth of the echo repeater is 40 m and is displayed with a black line in each image. The depth of the bottom is also shown as a black line at 92 m.

dependent fading over the 500 Hz bandwidth. Thus the leakage occurs in some sub-bands; the smaller and perhaps disjoint frequency support causes the dispersion of the leakage images. Finally, the target strengths of the targets are equal in this simulation, but the bottom image has a greater amplitude than the target. This suggests that the primary reason for the lack of reverberation in the data is a low bottom backscattering strength.

IV. CONCLUDING REMARKS

This paper presents a demonstration of the DORT technique with a system mounted on a ship in shallow water. The demonstration used linear frequency modulated (LFM) signals and time domain multiplexing (TDM) to develop a set of echoes from different source ensonifications. The use of these signals permitted the measurement of the decay time of each signal and the clear separation and construction of the data matrix. In a more realistic system, the development of suitable source signal multiplexing is needed in order to achieve acceptable interference effects. The TDM signals used in this study create interferences at earlier and later times [Gaumond *et al.* 2004]. Other multiplexing techniques, such as code division multiplexing (CDM), would create different interferences at earlier and later times. Intuitively, a

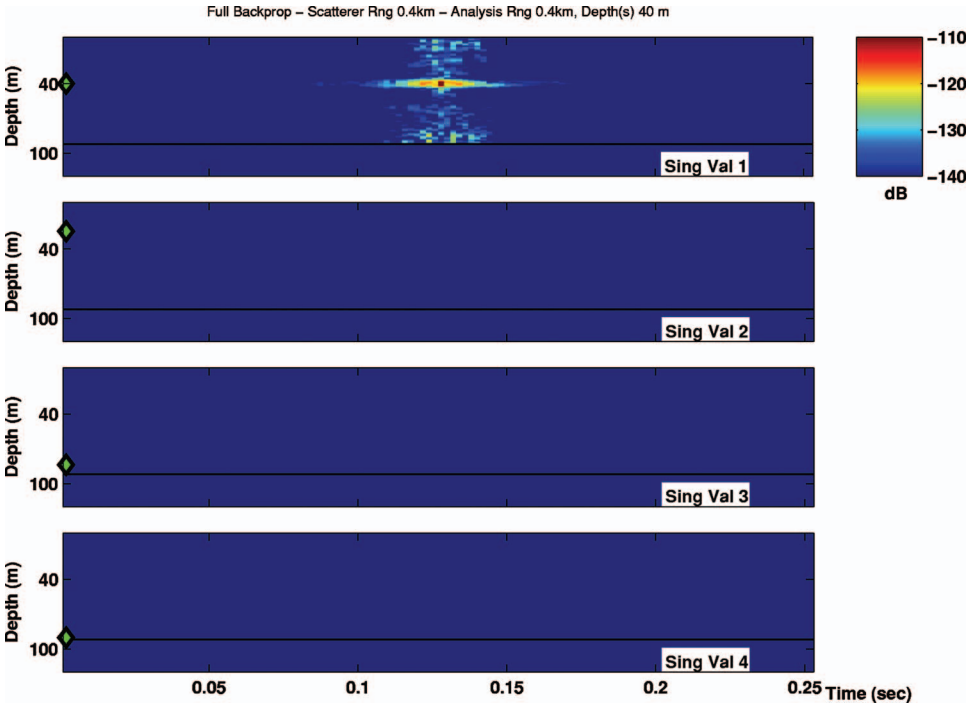


FIG. 11. The back-propagation images using noise-free, numerically simulated echoes that exactly correspond to the back-propagation Green's functions. The dynamic range of each image is 30 dB. The lower three panels have the same scale as the top panel.

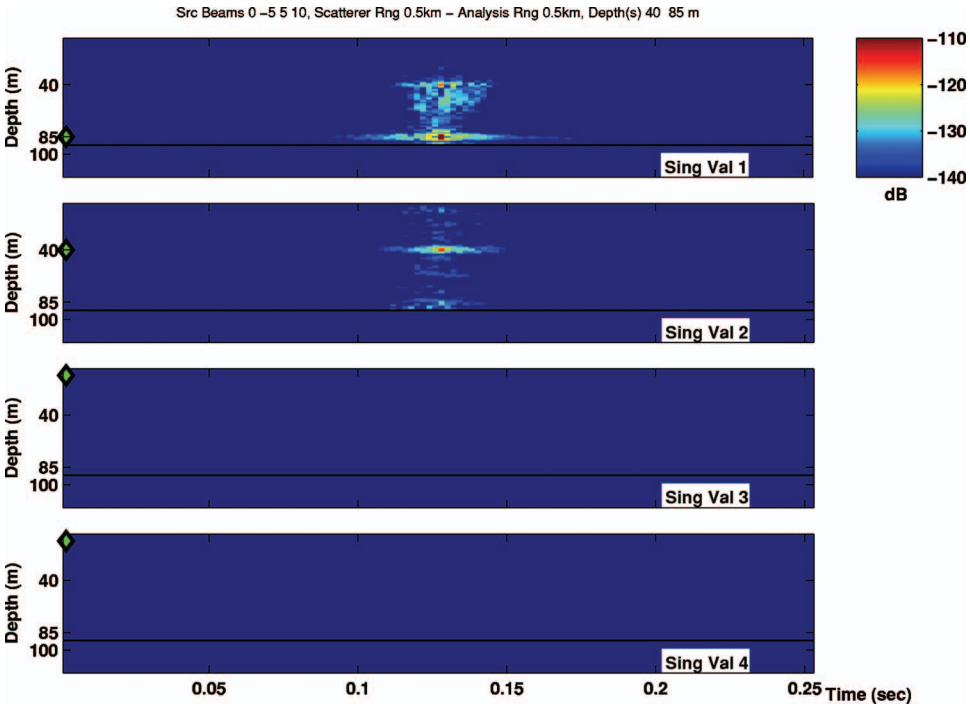


FIG. 12. The back-propagation images using noise-free, numerically simulated echoes from two identical scatterers, one at 40 m and the other at 85 m. The dynamic range of each panel is 30 dB and the scales are identical. The two signals are separated to a high degree, but not perfectly.

CDM technique should be preferable due to the more noise-like side lobes as opposed to the peaks found in the side lobes of the TDM signals. The study of various multiplexing techniques to preserve time invariance for moving sonars is called for in the bottom-reverberation limited case [Folegot *et al.* 2004].

The paper presents a derivation of expected results for the DORT processing of a data matrix comprised of a noiseless signal, pure noise, and also a mixture of a single target and noise. Not surprisingly the noise is separated into unequal amounts in the singular values. There is a benefit to using a higher rank matrix in the noise-limited case, but the tradeoff between using more source transmissions and the lowering of the noise level in the highest singular value needs to be studied in further detail.

Even though the SNR in our data is very high, the back-propagation images used time-reversal processing instead of inverse filtering [Tanter *et al.* 2000] because time-reversal processing is similar to matched filtering against noise in the low SNR limit. Back-propagation images using inverse filtering have lower side lobes and have been discussed for isolating a target in reverberation [Folegot *et al.* 2003].

This paper presents the limitations on coherence as a function of frequency. The DORT technique requires coherence between the sequence of signal transmissions. Even though the transmission sequence had a duration of only one second, the motion of the array still caused a loss of coherence at 3.25 kHz. The limitations of time invariance have been studied (Kuperman *et al.* 1997; Sabra and Dowling 2004). Our finding of short time invariance is probably due to the mounting of our equipment on heaving ships. Greater invariance, broken only by the motion of the ocean itself, could probably be obtained by bottom mounting or motion-isolated mounting to a ship or surface buoy. The coherent averaging technique of [Derode *et al.* 2001] overcomes some types of variability. Although this technique has some potential for overcoming some oceanic variability, the drifting of the echo repeater violates the requirements of this method.

Lastly, in Sec. II C the temporal response of $\sigma(\nu,t)$ is related to the propagation and target spreading. This suggests that information encoded into the temporal response of $\sigma(\nu,t)$ is potentially useful for classification.

ACKNOWLEDGMENTS

The authors would like to thank the captains and crews of the R/V Cape Henlopen and R/V Endeavor for their excellent work and support during our cruise. The authors would also like to acknowledge discussions over the years with Dr. Claire Prada, Dr. Matthias Fink, Dr. Thomas Folegot, and LT Alan Meyer (USNR). This research was supported by the Office of Naval Research.

Burdic, W. S. (**1991**). *Underwater Acoustic Systems Analysis*, 2nd ed., Prentice-Hall, New Jersey.

Carin, L., Liu, H., Yoder, T., Couchman, L., Houston, B., and Bucaro, J. (**2004**). "Wideband time-reversal imaging of an elastic target in an acoustic waveguide," J. Acoust. Soc. Am. **115**, 259–268.

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

Derode, A., Tourin, A., and Fink, M. (**2001**). "Random multiple scattering of ultrasound. I. Coherent and ballistic waves," Phys. Rev. E **64**, 036605-1–036605-7.

Folegot, T., Kuperman, W. A., Song, H. C., Akal, T., and Stevenson, M. (**2004**). "Using acoustic orthogonal signal in shallow water time-reversal applications," J. Acoust. Soc. Am. **115**, 2468.

Folegot, T., Prada, C., and Fink, M. (**2003**). "Resolution enhancement and separation of reverberation data from target echo with the time reversal operator decomposition," J. Acoust. Soc. Am. **113**, 3155–3160.

Gaumond, C. F, Fromm, D. M, Lingevitch, J., Menis, R., Edelmann, G., and Kim, E. (**2004**). "Application of DORT to active sonar," Conference Proceedings of Oceans '04, OTO'04 Conference Committee (IEEE, Piscataway, NJ).

Gauss, R. C. (2003). (Private communication).

Goff, J. A., Swift, D. J. P., Duncan, C. S., Mayer, L. A., and Hughes-Clarke, J. (**1999**). "High resolution swath sonar investigation of sand ridge, dune, and ribbon morphology in the offshore environment of the New Jersey

Margin," Mar. Geol. **161**, 309–339.
Golub, G. H., and Van Loan, C. F. (**1991**). *Matrix Computations* (The Johns Hopkins University Press, Baltimore), p. 71.
Gradstein, I. S., and Rizikh, I. M. (**1971**). *Table of Integrals, Sums, Sequences and Products* (Science Publishers, Moscow), (in Russian), p. 44.
Kuperman, W. A., Hodgkiss, W. S., Song, H. C., Akal, T., Ferla, C., and Jackson, D. R. (**1997**). "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40.
Lingevitch, J. F., Song, H. C., and Kuperman, W. A. (**2002**). "Time reversed reverberation focusing in a waveguide," J. Acoust. Soc. Am. **111**, 2609–2614.
Mast, T. D., Nachman, A. I., and Waag, R. C. (**1997**). "Focusing and imaging using eigenfunctions of the scattering operator," J. Acoust. Soc. Am. **102**, 715–725.
Mehta, M. (**1991**). *Random Matrices* (Academic Press, Boston, MA).
Montaldo, G., Tanter, M., and Fink, M. (**2001**). "Real time identification of several targets by iterative time reversal," J. Acoust. Soc. Am. **112**, 2308–2309.
Prada, C., Manneville, D., Spoliansky, D., and Fink, M. (**1996**). "Decomposition of the time reversal operator: Detection and selective focusing on two scatterers," J. Acoust. Soc. Am. **99**, 2067–2076.
Prada, C., Thomas, J.-L., and Fink, M. (**1995**). "The iterative time reversal process: Analysis of the convergence," J. Acoust. Soc. Am. **97**, 62–71.
Roux, P., Derode, A., Peyre, A., Tourin, A., and Fink, M. (**2000**). "Acoustical imaging through a multiple scattering medium using a time-reversal mirror," J. Acoust. Soc. Am. **107**, L7–L12.
Sabra, K., and Dowling, D. (**2004**). "Broadband performance of a time reversing array with a moving source," J. Acoust. Soc. Am. **115**, 2807–2817.
Tanter, M., Thomas, J.-L., and Fink, M. (**2000**). "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**, 223–234.

Mechanism for bandpass frequency characteristic in distortion product otoacoustic emission generation

P. F. Fahey[a)]
Department of Physics/Electrical Engineering, University of Scranton, Scranton, Pennsylvania 18510

B. B. Stagner and G. K. Martin
Jerry Pettis Memorial Veterans Medical Center, Loma Linda, California 92357

(Received 24 August 2005; revised 25 October 2005; accepted 7 November 2005)

It is commonly observed that the levels of the $2f_1$-f_2 and the other mf_1-nf_2 ($m=n+1=$integer) distortion product otoacoustic emissions (DPOAEs) initially increase in level for fixed f_2 as $f_1 \rightarrow f_2$, starting at $f_1 \ll f_2$, and then begin to decrease. When these DPOAE levels are plotted as a function of either the distortion product frequency (f_{dp}) or f_2/f_1 the curve has an approximate bandpass shape. It has been hypothesized that this effect is due to (1) a second filter, (2) suppression of distortion generation by the primary tones, (3) reemission of distortion products from the distortion product place (on the basilar membrane), (4) the presence of an even order nonlinearity, and (5) cancellation of the DPOAE due to the vector addition of multiple sources of distortion product. In this study distortion products were produced with f_1 sweep conditions where there would be minimal vector cancellation of multiple DPOAE sources. It was observed that under this condition, there is no or minimal bandpass shape of the DPOAE generation curve. Therefore, the data support the hypothesis that the bandpass shape obtained with traditional f_1 sweeps is due to vector cancellation from multiple sources. [DOI: 10.1121/1.2146088]

PACS number(s): 43.64.Jb, 43.64.Bt, 43.64.Ri, 43.64.Kc [WPS] Pages: 991–996

I. INTRODUCTION

One well-known and well-documented property of the distortion product otoacoustic emissions (DPOAEs) observed in the mammalian ear canal is that (at moderate input levels of two primary tones whose frequencies are f_1 and f_2 and whose levels are L_1 and L_2) the levels of the $2f_1$-f_2 DPOAE and $3f_1$-$2f_2$ DPOAE (indeed, any DPOAE of the form mf_1-nf_2, where $m=n+1$) decrease both in the limit as $f_1 \rightarrow f_2$ and as $f_1 \rightarrow f_2/2$. See Figs. 1 and 2. To explain this feature of the DPOAE data, it has been hypothesized that there is (1) a second filter (Brown and Gaskill, 1990; Brown *et al.*, 1992; Allen and Fahey, 1993); (2) suppression of the distortion products by the primary tones (Kanis and deBoer, 1997); (3) cancellation due to reemission from the distortion product place (Neely and Stover, 1997); (4) the presence of even order distortion that produces nulling of the DPOAE (Lukashkin and Russell, 2001); and (5) cancellation of the DPOAE due to multiple sources of distortion that add vectorially in the ear canal (Neely and Stover, 1997; Talmadge *et al.*, 1998; van Hengel and Duifhuis, 2000; Shera, 2002).

The experiment described here was designed to test the hypothesis that the bandpass shape of DPOAE generation is due to vector cancellation from multiple sources. Basilar membrane (BM) response (at a given place) to a sweep of a single tone is characterized by a slowly increasing level response accompanied by a slowly decreasing phase response until the frequency approaches the best frequency (BF) of that place (Robles and Ruggero, 2001). Near the BF both the level and phase decrease rapidly. For physical reasons, DPOAEs must be generated at or near the f_2 place (because here is where the mixing of both f_1 and f_2 in the nonlinearity will be greatest). Consider the $2f_1$-f_2 DPOAE. As $f_1 \rightarrow f_2$ the phase of f_1 near the f_2 place as a function of position begins to decrease rapidly; therefore, the local phase of the $2f_1$-f_2 DPOAE at generation will change twice as rapidly as the f_1 phase at the site of the nonlinearity. Moreover, the distortion product signal also acquires phase appropriate to the BM region where it is being generated because it is effectively a signal that is driving the BM at the place of generation. If a component of the DPOAE is generated at a place x_i on the BM, then it will carry a phase equal to $2\varphi_1(x_i)-\varphi_2(x_i)+\varphi_{dp}(x_i)$, where $\varphi_1(x_i)$ is the phase of the excitation at frequency f_1 at BM location x_i relative to the ear canal phase of f_1, etc. (Shera, 2002). [For the general mf_1-nf_2 DPOAE, the corresponding phase would be $m\varphi_1(x_i)-n\varphi_2(x_i)+\varphi_{dp}(x_i)$.] Assuming that the DPOAE observed in the ear canal is the vector sum of individual DPOAEs generated at many places, x_i, in the vicinity of f_2, cancellation would be expected to occur as $f_1 \rightarrow f_2$ because the vectors in the vector sum have increasing phase rotations relative to one another. As $f_1 \rightarrow f_2$, $f_{dp} \rightarrow f_1$ so that, to a first approximation, the relative phases between the adjacent sources is $2\varphi_1(x_{i+1})-2\varphi_1(x_i)$ (Shera, 2002) and this is changing rapidly. The net result of the rapid spatial change in phases for the set of spatially distributed sources will be a partial vector cancellation of these sources in the ear canal (the point of observation). The decrease in the DPOAE level as $f_1 \rightarrow f_2$ is due to wave interference.

[a)]Electronic mail: faheyp1@scranton.edu

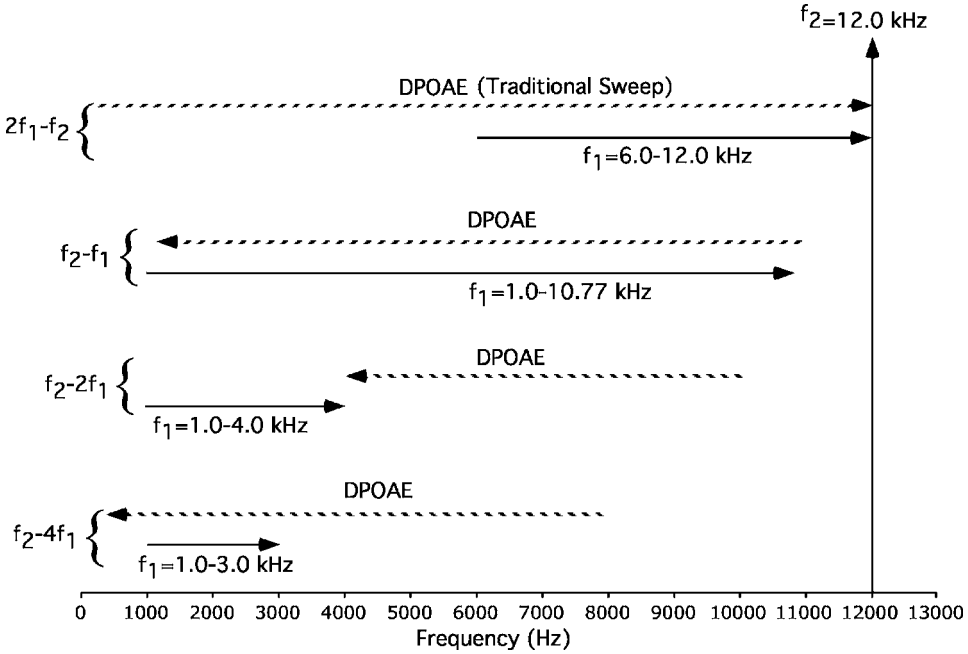


FIG. 1. Representation of the frequency ranges of a traditional DPOAE and the nontraditional DPOAEs and the frequency range of f_1 for fixed f_2 = 12 kHz.

Why does the DPOAE level decrease as $f_1 \rightarrow f_2/2$, that is, as $f_{dp} \rightarrow 0$? This is not due to rapidly changing phase but rather because the level of f_1 tone at the site of mixing is decreasing and, also eventually, because the low-frequency DP signal is not efficiently reverse transmitted to the ear canal. Therefore, a basilar membrane without a second filter will show a bandpass response for DPOAE generation (van Hengyl and Duifhuis, 2000; Neely and Stover, 1997; Kanis and deBoer, 1997; Matthews and Molnar, 1986). For a fixed value of f_2 let the value of the frequency where the DPOAE is a maximum be labeled as f_z. The second filter was proposed to act in concert with this mechanism to explain some DPOAE data where f_z seemed to be independent of DPOAE order. If it were possible to construct an experiment where the first filter that is due to vector addition of multiple sources with rapidly changing phase was rendered either inoperative or minimal, then the second filter should be laid bare and be easily detected.

Here we have generated DPOAEs under conditions where the f_1 tone is not undergoing rapid phase change at the site of generation because $f_1 \ll f_2$. The DPOAEs studied under this condition were f_2-f_1, f_2-$2f_1$, f_2-$3f_1$, and f_2-$4f_1$ (e.g., with f_2=12 kHz and 1 kHz$<f_1<$5 kHz, the f_2-$2f_1$ DPOAE has a frequency 4 kHz$<f_{dp}<$10 kHz and at the 12 kHz place f_z would be expected to be at around 8 or 9 kHz. Thus, the f_2-$2f_1$ DPOAE should have a bandpass shape centered at or near these frequencies.

This is the first study, to our knowledge, to systematically employ f_2-$2f_1$ and f_2-$4f_1$ DPOAEs to probe cochlear response. Figure 1 shows the range of the DPOAE frequencies and the f_1 frequencies for an f_2=12 kHz for the traditional $2f_1$-f_2 DPOAE and the set of nontraditional DPOAEs used in this study.

II. METHODS

A. Experimental animals

Subjects were two young, adult albino rabbits weighing between 3 and 4 kg. The DPOAE measurements were carried out in a sound-proofed booth, with the awake rabbit secured in a standard plastic restrainer, and its head supported in position by a Velcro chin strap fixed to the restrainer. The experimental-animal protocol was reviewed and approved locally by the Institutional Animal Care and Use Committee of the University of Colorado Health Sciences Center.

B. DPOAE measurement procedures

As an initial screening procedure, DP-grams, i.e., DPOAE level as a function of frequency, with primary-tone levels held constant were collected prior to other emission measurements described below. The DP-grams were collected at geometric-mean (GM) frequencies [GM=$(f_1 \times f_2)^{0.5}$], in 0.1-oct steps, from 1.414 to 18.37 kHz (f_2 =1.581–20.549 kHz). The primary tones (f_2/f_1=1.25) were either equilevel at $L_1=L_2$=45, 50, 55, 60, or 65 dB SPL, or with L_2 lower than L_1 by an amount that depended on the level of L_1, i.e., L_1-L_2=50–35, 55–45, and 60–55 dB SPL, which had previously been determined from a $2f_1$-f_2 DPOAE growth function that maximized sensitivity to cochlear insult (Whitehead *et al.*, 1995). These procedures have been previously described in detail elsewhere (Whitehead *et al.*, 1995). Briefly, the f_1 and f_2 primary tones were generated by a 16-bit digital signal processing (DSP) board (Digidesign, Audiomedia), mounted in a Macintosh Quadra personal microcomputer, and presented over ER-2 (Etymotic Research) loudspeakers. Ear-canal sound pressure, which was measured by an ER-10A (Etymotic Research) microphone, was sampled and synchronously averaged (n=4) by the DSP board. From a 2048-point fast Fourier transform (FFT) of the 4096-point time sample, the levels of the $2f_1$-f_2 DPOAE, and the related noise floor, were extracted. Both rabbits exhibited $2f_1$-f_2 emissions within normal limits bilaterally as compared to our laboratory database of rabbit emissions. The range of f_1 frequencies was from just below f_2 to 1 kHz. Because f_2-f_1, f_2-$2f_1$, f_2-$3f_1$, and f_2-$4f_1$ de-

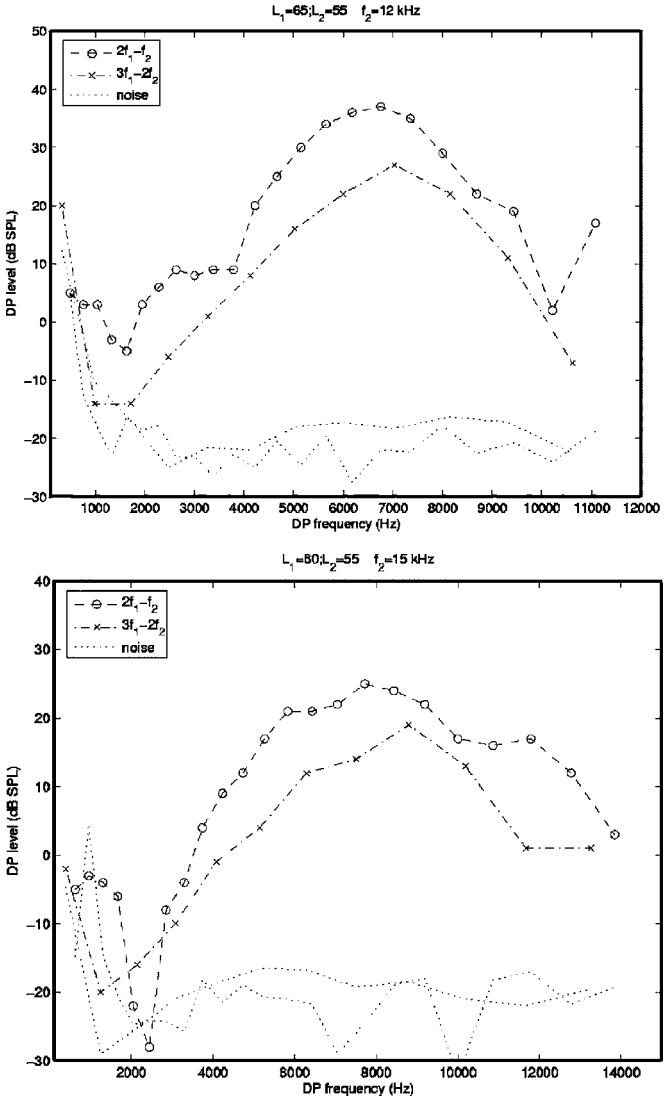


FIG. 2. The levels (in dB SPL) of the $2f_1$-f_2 and $3f_1$-$2f_2$ DPOAEs are plotted as a function of DPOAE frequency for $f_2=15$ kHz (right) and also $f_2=12$ kHz (left). The levels of f_1 are $L_1=65$ dB SPL (left) and $L_1=60$ dB SPL (right) and the level of f_2 is $L_2=55$ dB SPL. When $f_2=12$ kHz, the $2f_1$-f_2 is maximum at about 7.0 kHz, so $f_2/f_1=1.26$. When $f_2=15$ kHz, again $2f_1$-f_2 is maximum at about 8.0 kHz, so now $f_2/f_1=1.30$. The noise floors are plotted without symbols and are generally at −20 dB SPL or less (except at low frequencies).

crease in frequency as f_1 increases (moves toward f_2) and because f_1, $2f_1$, $3f_1$, $2f_1$-f_2, and $3f_1$-$2f_2$ increase in frequency, care must be taken that none of these latter five signals contaminate any of the four signals used as probes in this study. Hence, the raw data were processed to exclude all instances where the targeted DPOAE was possibly contaminated with another DPOAE (or one of the primary tones) that happened to have the same frequency.

C. L_1-L_2 swept f_1 spectral databases

Complete 2048 point spectra (0–22 kHz) of level and phase were obtained from synchronously averaged time samples ($n=16$) for each possible combination of L_1 and L_2 from 80 to 35 dB SPL in 5-dB steps with f_2 held constant at either 12 or 15 kHz to produce an L_1-L_2 DPOAE response area. L_1-L_2 response areas were collected at each f_2/f_1 ratio as f_1 was swept in various step sizes depending upon the DPOAE of interest. To elicit the typical "bandpass-like" ratio functions of $2f_1$-f_2 and $3f_1$-$2f_2$, f_1 was swept in 0.04 step sizes of f_2/f_1 ratio from 6 to 12 kHz for $f_2=12$ kHz and from 7.5 to 15 kHz for $f_2=15$ kHz. In order to obtain similar functions for f_2-$2f_1$ and f_2-$4f_1$, f_1 was swept in $-\frac{1}{12}$-oct steps from 4 to 1 kHz for both $f_2=12$ kHz and $f_2=15$ kHz. Finally, in order to obtain similar data for f_2-f_1, f_1 was swept from 10.77 to 1 kHz in $-\frac{1}{7}$-oct steps for $f_2=12$ kHz and from 14.25 to 0.9 kHz in $-\frac{1}{6}$-oct step sizes for $f_2=15$ kHz. All of these sweeps were repeated in a 1.5-cc hard-walled cavity and emission artifacts were noted to be generally below −10 dB SPL in level. Because of the time-consuming nature of the protocol all $f_2=12$ kHz experiments were performed on the left ear of one rabbit while the remaining $f_2=15$ kHz runs were performed on the left ear of the other rabbit.

The resulting 2500 spectra for each f_1 sweep were databased in Microsoft Excel X for Macintosh and evaluated for the presence of different emissions: For instance, f_2-$3f_1$ was evaluated and often found not to be present at significant level above the noise floor. For the emissions that were of sufficient level and of interest for this study, L_1 vs. L_2 DPOAE level response areas for each unique f_2/f_1 ratio were extracted from the database and plotted as contours. DPOAE level versus emission frequency for a given L_1-L_2 level were extracted and plotted as line graphs to produce the familiar f_2/f_1 ratio functions. This allowed for a detailed comparison of various emission frequency versus emission level functions at all the possible L_1-L_2 combinations that were presented. The optimal L_1-L_2 level combinations for eliciting each emission were identified, preliminarily evaluated for possible "bandpass-like" characteristics, and employed in the final experiments described below. The probe microphone placement was such that there were no or minimal standing waves observed in the ear canal response.

III. RESULTS

A. The phenomenon to be explained—the classical result

Figure 2 shows the classical bandpass type of DPOAE frequency generation for rabbits at $f_2=12$ and 15 kHz. The value of f_z is at about 7 kHz when $f_2=12$ kHz and is at about 8 kHz when $f_2=15$ kHz. This is almost a whole octave below the f_2 tone. Generally, in guinea pigs at low levels (Brown and Gaskill, 1990) and humans, f_z is about $\frac{1}{2}$ oct below f_2 (Brown and Gaskill, 1990; Stover *et al.*, 1999). In cats there is a frequency map that is fairly independent of the level of f_2 (Allen and Fahey, 1993). Harris *et al.* (1989) and Stover *et al.* (1999) show data that suggest that a similar map in humans would be dependent on both the level and the frequency of f_2. The ratio of f_z/f_2 in rabbits is on the low side of the commonly used experimental animals. It is also observed in rabbits that the DPOAE bandpass characteristic is level labile at the f_2 frequencies used here. When the level of f_2, L_2, is greater than 75 dB SPL the bandpass behavior is no longer evident. It should also be mentioned that there was generally little evidence (from the phase data not shown

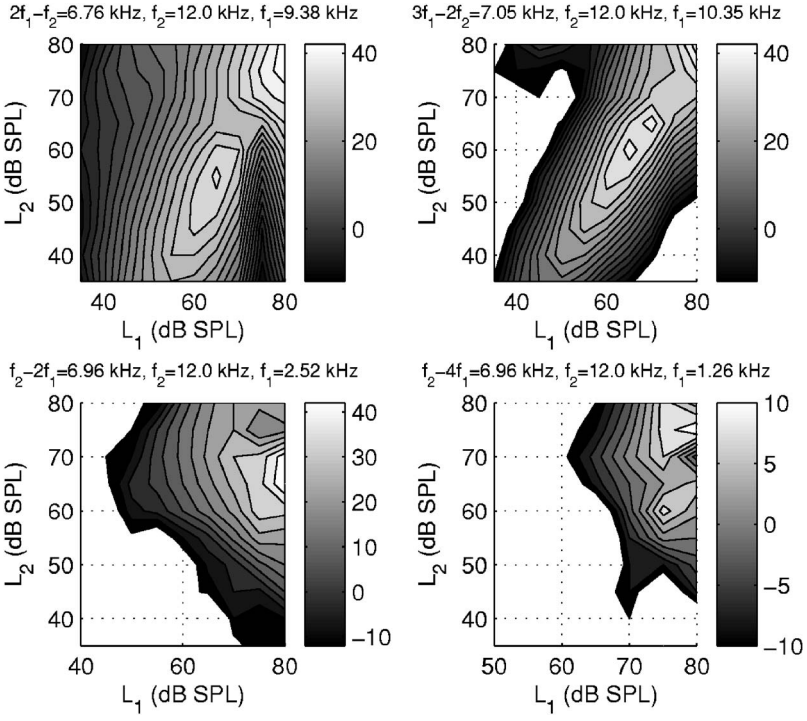


FIG. 3. For constant f_2, f_1, and f_{dp} the dependence of DPOAE level on the levels, L_1 and L_2, of the primary tones is displayed for $f_{dp}=2f_1-f_2$ (top left), $f_{dp}=3f_1-2f_2$ (top right), $f_{dp}=f_2-2f_1$ (bottom left), and $f_{dp}=f_2-4f_1$ (bottom right).

here) of a DPOAE component that was reemitted from the DP place for the $2f_1$-f_2 or $3f_1$-f_2 DPOAEs except at the smallest ratios of f_2/f_1 (Knight and Kemp, 2001) and never was there any evidence in the region where these two DPOAEs are at maximum level. Hence, the presence of a reemitted component is not likely to be affecting the results and conclusions presented here. Figure 3 shows how four of the DPOAEs studied here depend upon the levels of the primary tones. Notice that there is much less dynamic range available with the f_2-$2f_1$ and f_2-$4f_1$ DPOAEs than with the traditional DPOAEs, especially considering that we wish to use $L_2<75$ dB SPL.

B. New (nonclassical) DPOAE frequency response in rabbits

Figure 4 shows the f_2-f_1, f_2-$2f_1$, and f_2-$4f_1$ DPOAE responses for $f_2=15$ kHz and 12 kHz and also the f_2-$3f_1$ DPOAE for $f_2=12$ kHz. It is evident that there is no bandpass frequency response like there is in the $2f_1$-f_2 and $3f_1$-$2f_2$ DPOAEs. If there were a second filter that was responsible for some of the bandpass nature of the $2f_1$-f_2 and $3f_1$-$2f_2$ DPOAEs, then one would expect evidence of this filter in the f_2-f_1, f_2-$2f_1$, f_2-$3f_1$, and f_2-$4f_1$ DPOAEs. Furthermore, none of the other proposed mechanisms seems to be working to produce a bandpass frequency response. Because, in the overwhelming majority of our data sets, there is no or minimal bandpass character to the level versus frequency response, we conclude that there is no second filter in rabbit DPOAE generation nor are the other proposed mechanisms operative. And, more strongly, we conclude that the mechanism responsible for the bandpass character of the mf_1-nf_2 ($m=n+1$) DPOAEs is wave interference due to the vectorial summation of sources with rapidly changing phases. When the rapid phase rotation of the multiple DPOAE sources is disabled, the bandpass frequency response is disabled.

The latencies (calculated from the gradient of the phase with angular frequency) averaged the same for all of the DPOAEs at a given value of f_2. That value was consistent with the delay expected from the f_2 place. It should also be noted that there is considerable variability in the data of Fig. 4. Data more recently obtained at a much finer frequency resolution than shown in Figs. 2 and 4 show evidence of wave interference. Of course this implies at least a second source of the DPOAE. The most likely second source would be a reemission from the distortion product place. However, the data so far are not always consistent with this simple picture. Hence, the possible source of the second wave remains under study. But, certainly, wave interference underlies much of the variability of the data in Fig. 4. The higher frequency resolution data and the lower frequency resolution data both support the conclusion that the bandpass frequency response of the traditional DPOAEs is due to vector cancellation upon generation.

IV. DISCUSSION

A. Source of the bandpass character of the $2f_1$-f_2 and $3f_1$-$2f_2$ DPOAEs

The model of Matthews and Molnar (1986) showed a bandpass character to DPOAE generation in a nonlinear basilar membrane (BM) model that had no second filter. Later nonlinear BM models, such as the ones of Neely and Stover (1997), van Hengyl and Duifhuis (2000), and Kanis and de-Boer (1997) that also had no second filter, also showed a bandpass characteristic of the DPOAE generation. Each of these models had a rapidly changing phase in the vicinity of the BM characteristic place. Therefore, each of these models

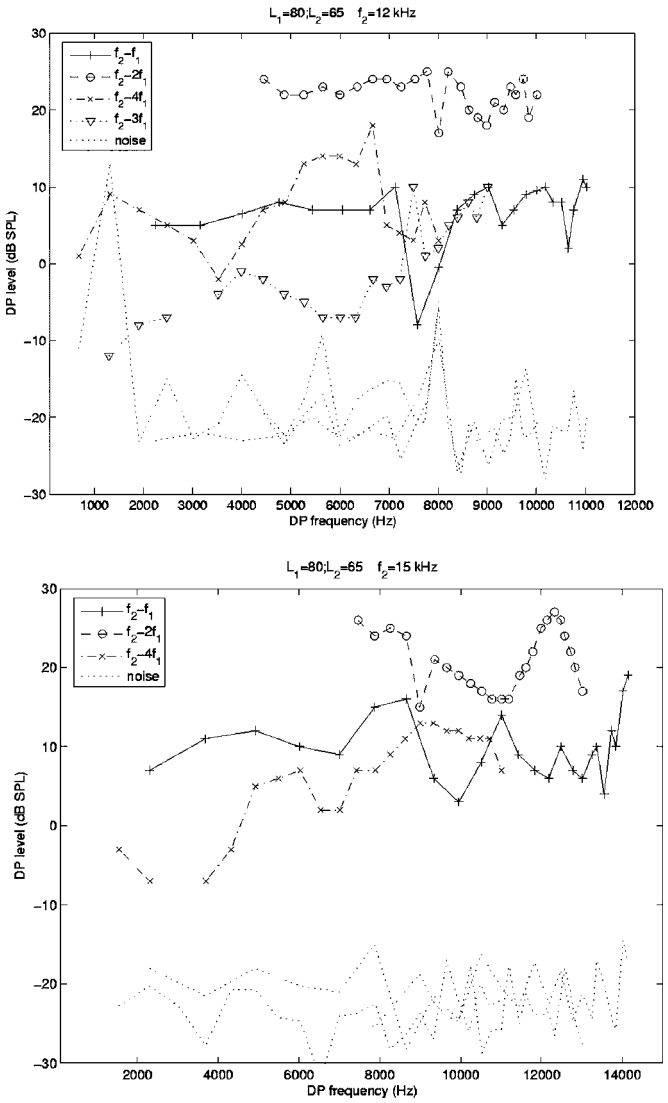


FIG. 4. The levels (in dB SPL) of the f_2-f_1, f_2-$2f_1$, and f_2-$4f_1$ DPOAEs are plotted as a function of DPOAE frequency for $f_2=15$ kHz (right) and $f_2=12$ kHz (left). Also, the f_2-$3f_1$ DPOAE is shown for $f_2=12$ kHz. The level of f_1 is $L_1=80$ dB SPL and the levels of f_2 are $L_2=60$ (left) and 65 (right) dB SPL. Two of the three noise floors are plotted without symbols. All of the noise floors were essentially the same and at or below −20 dB SPL over the frequency range of interest.

would be expected to exhibit a DPOAE bandpass character due to the rapidly changing phase of the excitation at the f_2 place as $f_1 \rightarrow f_2$. Assuming that DPs are generated over a region of BM where both the f_1 and the f_2 excitation are large enough to mix in a distributed nonlinearity, then the DPOAE observed in the ear canal will be a vector sum of a number of distributed sources. These sources will have varying phases when $f_1 \rightarrow f_2$ and wave cancellation in the ear canal would be expected, thereby reducing the level. [Alternative explanations for this bandpass effect have been proposed. Neely and Stover (1997) have suggested that some of the bandpass characteristic is due to reemission of the distortion product from it characteristic place and Kanis and deBoer (1997) postulate that it is due to suppression of the distortion products by the primary tones.]

B. Turning off the vector cancellation

By using f_2-f_1, f_2-$2f_1$, f_2-$3f_1$, and f_2-$4f_1$ DPOAEs, where $f_1 \ll f_2$, one can generate distortion products at and near the f_2 place under conditions where the phase of f_1 changes very little as f_1 itself changes. Therefore, the role of f_1 phase in vector cancellation is turned off and vector cancellation is markedly turned down. Now, with the "first filter" largely disabled it should be possible to see the "second filter" very cleanly and also to see evidence of any other mechanism that decreases the level of the DPOAE as $f_{dp} \rightarrow f_2$. In the data reported here, there is neither evidence of a second filter nor evidence of any of the other hypothesized nonvector cancellation mechanisms.

V. CONCLUSION

We find no evidence of a second filter playing a role in DPOAE generation in rabbit and we doubt that rabbit DPOAE generation differs from other experimental animals. However, this experiment should be repeated in other species in order to see it if it can be generally concluded that there is no role for a second filter in DPOAE generation in general. The results here support the hypothesis that the amplitude of the $2f_1$-f_2 and $3f_1$-$2f_2$ DPOAEs as $f_2/f_1 \rightarrow 1$ is affected by vector cancellation from multiple generation sites (Neely and Stover, 1997; Talmadge *et al.*, 1998; van Hengel and Duifhuis, 2000; Shera, 2002). These experiments also showed no evidence of suppression of the distortion products by the primaries as proposed by Kanis and deBoer (1997) or by reemission from the distortion product place as proposed by Neely and Stover (1997). Neither was there any evidence of the mechanism proposed by Lukashkin and Russell (2001). It is concluded that the bandpass characteristic of lower sideband DPOAE frequency responses is likely due to wave interference (cancellation) by multiple DPOAE sources.

ACKNOWLEDGMENTS

This research was supported in part by funds from the Public Health Service [supported in part by the Public Health Service (DC000613, DC003114) and the Research Fund of the Department of Otolaryngology, University of Colorado Health Sciences Center, Denver CO] and DOE (PFF). We thank two anonymous reviewers for many useful comments and suggestions.

Allen, J. B., and Fahey, P. F. (**1993**). "A second cochlear-frequency map that correlates distortion product and neural tuning mechanisms," J. Acoust. Soc. Am. **94**, 809–816.

Brown, A. M., and Gaskill, S. A. (**1990**). "Can basilar membrane tuning be inferred from distortion measurement?," in *The Mechanics and Biophysics of Hearing*, edited by P. Dallos, C. D. Geisler, J. W. Mathews, M. A. Ruggero, and C. R. Steele (Springer-Verlag, New York), pp. 164–169.

Brown, A. M., Gaskill, S. A., and Williams, D. M. (**1992**). "Mechanical filtering of sound in the inner ear," Proc. R. Soc. London, Ser. B **250**, 29–34.

Harris, F. P., Lonsbury-Martin, B. L., Stagner, B. B., Coats, A. C., Martin, G. K., (**1989**). "Acoustic distortion productsd in humans: Systematic changes in amplitude as a function of f_2/f_1 ration", J. Acoust. Soc. Am. **85**, 220–229.

Kanis, L. J., and deBoer, E. (**1997**). "Frequency dependence of acoustic products in a locally active model of the cochlea," J. Acoust. Soc. Am. **101**, 1527–1531.

Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.

Lukashkin, A. N., and Russell, I. (**2001**). "Origin of the bell-like dependence of the DPOAE amplitude on primary frequency ratio," J. Acoust. Soc. Am. **110**, 3097–3106.

Mathews, J. W., and Molnar, C. E. (**1986**). "Modeling of intracochlear and ear canal distortion product $2f_1$-f_2," in *Peripheral Auditory Mechanisms*, edited by J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis (Springer-Verlag, New York), pp. 258–265.

Neely, S. T., and Stover, L. J. (**1997**). "A generation of distortion products in a model of cochlear mechanics," in *Diversity in Auditory Mechanics*, edited by E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, C. R. Steele, and E. Hecht-Poinar (World Scientific, Singapore), pp. 434–440.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Shera, C. A. (**2002**). "Wave interference in the generation of reflection- and distortion-source emissions," in *Biophysics of the Cochlea: from Molecules to Models*, edited by A. W. Gummer (World Scientific, Singapore), pp. 439–453.

Stover, L. J., Neely, S. T., and Gorga, M. P. (**1999**). "Cochlear generation of intermodulation distortion revealed by DPOAE frequency functions in normal and impaired ears," J. Acoust. Soc. Am. **106**, 2669–2678.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

van Hengel, P. W. J., and Duifhuis, H. (**2000**). "The generation of distortion products in a nonlinear transmission line model of the cochlea," in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 409–415.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears: I. Effects of decreasing L_2 below L_1," J. Acoust. Soc. Am. **97**, 2346–2358.

Invariance principles for cochlear mechanics: Hearing phases

H. M. Reimann[a)]
Institute of Mathematics, University of Berne, Sidlerstrasse 5, CH-3012 Berne, Switzerland

(Received 2 May 2005; revised 28 November 2005; accepted 28 November 2005)

A functional model of the cochlea is devised on the basis of the results from classical experiments. The basilar membrane filter is investigated in detail. Its phase is close to linear in the region around the peak of the amplification. On one side this has consequences for the time analysis and on the other side this has led to a prediction on phase perception for very simple combinations of tones, a prediction which is now confirmed by experiments. Equivariance under the dilation group permits one to describe the model by a wavelet transform [Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992)]. The wavelet is discussed in reference to the phase analysis of the basilar membrane filter. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159428]

PACS number(s): 43.64.Kc [NF] Pages: 997–1004

I. PHASE ANALYSIS

A. The basilar membrane filter

Acoustic signals which reach the ear are described in terms of pressure variations $f(t)$ of the air. In the outer ear the incoming pressure waves are channeled. The eardrum is set into motion and the middle ear passes this motion onto the oval window. At this place the media changes. Pressure waves of the cochlear fluid in the inner ear are generated. The fluid mechanical process in the inner ear has been described by Lighthill in his fundamental paper "Energy flow in the cochlea" in 1981. According to his theory there will be two superimposed waves, the slow and the fast wave. The fast wave travels along the entire cochlea at the speed of sound in a fluid. It has no direct effect on the basilar membrane. The slow wave arises from the pressure differences between the scala tympani and the scala vestibuli (the scala media and Reissner's membrane are neglected in this model). It is the slow wave that sets the cochlear partition into motion. Its properties are essentially determined by the impedance of the cochlear partition and by the amplification mechanism of the hair cells, both of which change their characteristics along the basilar membrane.

It is in the cochlea, the inner ear, where the acoustic signal is transformed into neural impulses. The inner hair cells are at the basis of this transduction. Bipolar nerve cells whose cell bodies lie within the spiral ganglion account for about 90% of the afferent innervation of the cochlea [see, e.g., Geisler (1998), Chap. 11]. These neurons are myelinated (therefore of type one) and their peripheral axons form a single synapse with an inner hair cell. The remaining afferent neurons are of type two. They are unmyelinated and their axons each form synapses with 10–20 clustered outer hair cells.

The transduction of the acoustic signal into nerve impulses is thus governed by the electromechanical properties of the hair cells and most of this translation is effectuated by the inner hair cells. In our present approach we tacitly assume that the nerve impulses directly depend on the deviation $u(x,t)$ of the basilar membrane from its resting position [i.e., $u(x,t)$ describes the movement of the cilia of the inner hair cells. These in turn trigger the nerve impulses]. The variable t denotes time and x is the distance from the stapes (oval window) measured in millimeters along the basilar membrane. Needless to say, our assumption is a gross simplification.

We thus study the process that associates to an incoming acoustic signal $f(t)$ the deviation $u(x,t)$ of the basilar membrane from its resting position.

With a specific view to the hearing process, there is however a difficulty: The amplitudes of the acoustic signals vary by many orders of magnitude (up to 100 dB) and the ear has to act compressively. The deviations $u(x,t)$ are at the size of a few nanometers and vary by only a few orders of magnitude. The requirement that a multiple $cf(t)$ of the signal $f(t)$ will lead to a multiple $cu(x,t)$ of the response can therefore hold only within rather narrow limits.

Obviously, hearing is time invariant. If we assume that the cochlear process is linear, then time invariance implies that the response to the special input signal

$$f(t) = \mathrm{Re}\, e^{i\omega t} = \cos \omega t, \quad \omega > 0, \tag{1}$$

has to be of the form

$$u(x,t) = \mathrm{Re}\{\hat{g}(x,\omega)e^{i\omega t}\}. \tag{2}$$

The function $\hat{g}(x,\omega)$ is called the basilar membrane filter. We express it in polar coordinates as

$$\hat{g}(x,\omega) = a(x,\omega)e^{i\vartheta(x,\omega)}. \tag{3}$$

In this paper we analyze the argument $\vartheta(x,\omega)$ of $\hat{g}(x,\omega)$. It describes the phase shift. The amplification is determined by the absolute value $a(x,\omega) \geqslant 0$ of $\hat{g}(x,\omega)$. This quantity will be analyzed in a subsequent paper.

Because of the compressive (nonlinear) behavior of the hearing process, the amplification factor $a(x,\omega)$ is sensitive to the amplitude A of the incoming signal $\mathrm{Re}\, Ae^{i\omega t}$. However the phase $\vartheta(x,\omega)$ remains largely unaffected (cf. the reference in Sec. I C to the work of Rhode and Recio, 2000). The

a)Electronic mail: reimann@math-stat.unibe.ch

cochlear process keeps exact track of the timing and acts compressively with respect to the sound level.

Assuming linearity, the effect of an arbitrary incoming signal $f(t)$ can be calculated from its Fourier decomposition:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega = \mathrm{Re} \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} \hat{f}(\omega) e^{i\omega t} d\omega \quad (4)$$

with

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt. \quad (5)$$

The superposition principle gives

$$u(x,t) = \mathrm{Re} \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} \hat{g}(x,\omega) \hat{f}(\omega) e^{i\omega t} d\omega. \quad (6)$$

B. The frequency position function

For a fixed position x along the basilar membrane the response to a sinusoidal input increases with the circular frequency up to a critical frequency ω_x, the so-called resonant frequency. If the frequency is increased beyond ω_x, the reaction quickly drops to zero. The function, which assigns to x the resonant frequency ω_x, is the frequency position function.

The inverse function is called the tonotopic axis. It assigns to a frequency ω the place x_ω on the basilar membrane for which ω is the resonant frequency. We will use the approximations

$$\omega_x = e^{(x^* - x)/\gamma}, \quad (7)$$

$$x_\omega = x^* - \gamma \log \omega \quad (8)$$

with

$$x^* = 35 \left(1 + \frac{\log(2\pi \cdot 20)}{\log 1000} \right), \quad (9)$$

$$\gamma = 35 \frac{1}{\log 1000}. \quad (10)$$

The constants are determined such that the basilar membrane $0 \leqslant x \leqslant 35$ covers the frequency range from 20 to 20 000 Hz—this range has been chosen for historic reasons, see, e.g., the survey by Greenwood (1990), the exact range is not so relevant for our consideration—

$$x_{2\pi \cdot 20} = 35 = x^* - \gamma \log(2\pi \cdot 20), \quad (11)$$

$$x_{2\pi \cdot 20\,000} = 0 = x^* - \gamma \log(2\pi \cdot 20\,000). \quad (12)$$

It should be noted at this point that the frequency position function can be described more exactly by introducing a shift of the frequency scale [Liberman (1982)]. To take this shift into account would make the mathematical analysis substantially more complicated. The resulting scale expansion affects low frequencies (<1 kHz) and it is clear that some modifications have to be made. For the sake of simplicity this is not done in this paper.

A word of caution about the definition of the tonotopic axis. If the frequency ω is fixed, then the absolute value $a(x,\omega)$ of the basilar membrane filter $\hat{g}(x,\omega)$ is an increasing function of x up to a certain point $x_{\max}(\omega)$ at which it attains its maximum. For $x > x_{\max}(\omega)$ the function $a(x,\omega)$ quickly drops to zero.

If $a(x,\omega)$ satisfies the invariance law

$$a(x - \gamma \log c, c\omega) = a(x,\omega) \quad (13)$$

[cf. equation (62)], then it follows that

$$x_{\max}(\omega) = x_\omega. \quad (14)$$

But otherwise, this need not necessarily be true.

C. Invariance

The pioneering *in vivo* experiments on the basilar membrane filter are due to Rhode (1971). These experiments were done on live squirrel monkeys and are based on the Mössbauer technique. Tiny emitters of gamma rays (Mössbauer sources) are placed at two different locations, one on the basilar membrane and one on the eardrum, at the place of contact with the handle of the malleus. Under a sinusoidal stimulation with frequency $\omega/2\pi$ the movement of the sources

$$u_0(t) = a_0 \cos(\omega t + \theta_0) \quad \text{at the eardrum} \quad (15)$$

and

$$u(x,t) = a(x) \cos(\omega t + \theta(x)) \quad \text{at } x \quad (16)$$

are compared. Using the Doppler effect this provides information on the ratio $a(x)/a_0$ of the amplitudes and on the phase difference $\theta - \theta_0$.

On the basis of Rhode's results Zweig (1976) postulated, [see also Lighthill (1981) p. 171] that the argument $\vartheta(x,\omega)$ of the basilar membrane filter depends on the ratio ω/ω_x only. With the expression

$$\omega_x = e^{(x^* - x)/\gamma} \quad (17)$$

for the frequency position function we obtain

$$\log \frac{\omega}{\omega_x} = \log \omega + \frac{x}{\gamma} - \frac{x^*}{\gamma}. \quad (18)$$

The postulate can thus be reformulated in the way that $\vartheta(x,\omega)$ is a function of $x + \gamma \log \omega$ only.

Expressed still in another way, we can say that the function $\vartheta(x,\omega)$ is invariant under dilations:

$$\vartheta(x,\omega) = \vartheta(x - \gamma \log c, c\omega), \quad c > o. \quad (19)$$

In fact, if one changes ω to $c\omega$, then in order to keep $x + \gamma \log \omega$ invariant one has to subtract $\gamma \log c$ from x.

Dilation invariance is a very important basic feature of the hearing process. This topic will be taken up again in the next chapter.

The experiments of Rhode (1971) further show that for fixed x the phase $\vartheta(x,\omega)$ is an affine function at least for values ω in a region which includes the interval

$$\tfrac{2}{3}\omega_x \leqslant \omega \leqslant \omega_x. \tag{20}$$

This is the essential range in which the amplification factor $a(x,\omega)$ is of a sizable magnitude. More recent investigations by Rhode and Recio (2000) on the phase of the basilar membrane filter in the chinchilla cochlea have confirmed that "At sufficiently low frequencies the input-output function becomes linear and the phase changes with level are absent" [Rhode and Recio (2000), p. 3321]. However at high characteristic frequencies (12 kHz for the given example), the phase changes somewhat with the intensity level [see Fig. 2(E) in Rhode and Recio]. The present section on the phase of the basilar membrane filter concentrates on aspects in the hearing process in which phase information is passed onto neuronal impulse rates. This becomes increasingly difficult for higher frequencies. Indeed the above-cited results show that this will not be feasible for high characteristic frequencies.

With these limitations in mind it is assumed that

$$\vartheta(x,\omega) = \kappa - \mu\frac{\omega}{\omega_x}. \tag{21}$$

Zweig's postulate then tells us that κ and μ are constants (they depend neither on x nor on ω).

A rough estimate based on Fig. 5 in Rhode (1971) gives

$$\mu \simeq 10\pi. \tag{22}$$

The consequences are that the deviation $u(x,t)$ of the basilar membrane which results from the incoming signal $f(t)$ is well approximated by the formula

$$u(x,t) = \mathrm{Re}\sqrt{\frac{2}{\pi}}\int_0^\infty a(x,\omega)e^{i\vartheta(x,\omega)}\hat{f}(\omega)e^{i\omega t}d\omega$$

$$\cong \mathrm{Re}\sqrt{\frac{2}{\pi}}\int_0^\infty a(x,\omega)e^{i\kappa}e^{i\omega(t-\mu/\omega_x)}\hat{f}(\omega)d\omega. \tag{23}$$

We stress that $a(x,\omega)$ is negligibly small for values (x,ω) for which $\vartheta(x,\omega)$ differs from $\kappa-\mu\omega/\omega_x$.

D. Time analysis

The introduction of a time delay,

$$\Delta t = \Delta t(x) = \frac{\mu}{\omega_x} = \mu e^{-(x^*-x)/\gamma}, \tag{24}$$

simplifies the analysis considerably. It should be noted, however, that this is a purely analytical tool. The time it takes for a cosine-wave with circular frequency ω to travel from the stapes $(x=0)$ to the frequency position x_ω is given by

$$\frac{1}{\omega}(\vartheta(0,\omega) - \vartheta(x_\omega,\omega)). \tag{25}$$

In our approximation

$$\vartheta(x_\omega,\omega) = \kappa - \mu. \tag{26}$$

Furthermore, for frequencies up to about 2000 Hz the value $\vartheta(0,\omega)$ is close to zero. This gives an effective travel time on the order of $(\mu-\kappa)/\omega$.

The effect of an incoming signal can now be written as

$$u(x,t+\Delta t) = \mathrm{Re}\sqrt{\frac{2}{\pi}}\int_0^\infty a(x,\omega)e^{i\kappa}\hat{f}(\omega)e^{i\omega t}d\omega$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{i\kappa\ \mathrm{sign}\ \omega}a(x,|\omega|)\hat{f}(\omega)e^{i\omega t}d\omega. \tag{27}$$

Let us first discuss the case $e^{i\kappa}=1$ (κ is an entire multiple of 2π). Then for every x the deviation $u(x,t+\Delta t)$ is obtained from f by filtering. The special feature is that the filter function $a(x,\omega)$ is real and non-negative. At a fixed point x all wave components arrive in phase. Only the amplitudes are subject to change. The slow wave travels along the cochlea as if there were no dispersion.

Recall that traveling waves are called dispersion free, if the travel speed is independent of the frequency. The slow wave would genuinely be dispersion free, if the equation $\vartheta(x,\omega)=-\mu\omega/\omega_x$ were true for the entire range.

As an example, if f is the Dirac function representing a click at time $t=0$, then $\hat{f}(\omega)=1/\sqrt{2\pi}$. Because of the cancellation effects in the integral, the function $u(x,t+\Delta t)$ (with x fixed) will then essentially be supported in a neighborhood of $t=0$. The impulse travels along the cochlea and will be around x at time $\Delta t(x)$.

If the constant κ is not an entire multiple of 2π, then the filter function $a(x,\omega)\exp(i\kappa\ \mathrm{sign}\ \omega)$ will change the phases $\arg\hat{f}(\omega)$ by the fixed constant κ $(\omega>o)$. The phase content is thus changed in a controlled way.

E. The action of the circle group

The circle group $S=\{e^{i\varphi}:\varphi\in\mathbb{R}\}$ acts on real valued signals. Its action is easily described on the Fourier transform side

$$L^2_{\mathrm{sym}} = \{h \in L^2(\mathbb{R}): h(\omega) = \overline{h(-\omega)}\}. \tag{28}$$

The functions $h\in L^2_{\mathrm{sym}}$ are the Fourier transforms of real valued functions in $L^2(\mathbb{R})$. The element $e^{i\vartheta}$ in the group S acts by

$$h(\omega) \to e^{i\vartheta\ \mathrm{sign}\ \omega}h(\omega). \tag{29}$$

The transformation $f(t)\to u(x,t+\Delta t)$ is determined by the filter function

$$e^{i\kappa\ \mathrm{sign}\ \omega}a(x,|\omega|). \tag{30}$$

It can be considered as the action of $e^{i\kappa}\in S$ followed by filtering with the non-negative filter function $a(x,\omega)$.

The value of the constant κ in our filter function is very much determined by the nature of the quantity under consideration. In our situation, $f(t)$ was defined to be the pressure of the acoustic signal. But on its way through the outer- and the middle-ear this signal undergoes some changes and it is not easy to determine the phase of the pressure of the fluid wave generated in the cochlea at the stapes, let alone to translate the phase of the pressure wave into the phase of the deviation $u(x,t)$ (which is being measured in Rhode's experiment).

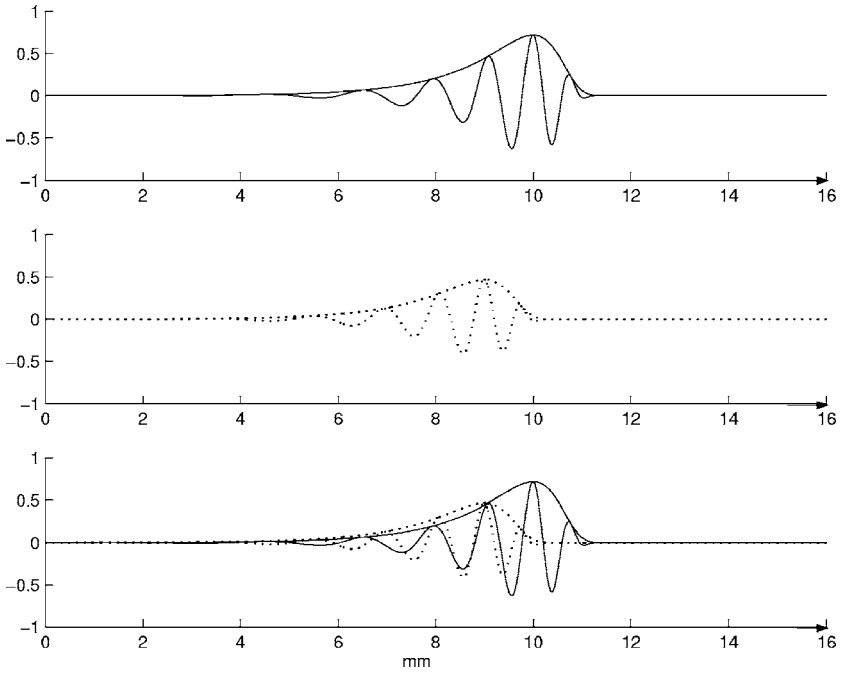


FIG. 1. Superposition of two harmonics along the basilar membrane, frequency ratio 2:3, relative scale.

Realistically, what can be expected is that κ varies very little with x. In examples 1 and 2 which follow, the value of κ is irrelevant. For example 3 we make the idealistic hypothesis that $e^{i\kappa}=1$.

F. Examples

1. Example 1: Two harmonics

The acoustic signal in this example consists of two harmonics with frequency ratio $k:j$. It is assumed that $1<k<j$ and that the integers k and j have no common divisor:

$$f(t) = \mathrm{Re}\{c_k e^{ik\omega t} + c_j e^{ij\omega t}\}. \tag{31}$$

The deviation along the cochlea is

$$u(x,t+\Delta t) = \mathrm{Re}(e^{i\kappa}\{c_k a(x,k\omega)e^{ik\omega t} + c_j a(x,j\omega)e^{ij\omega t}\}). \tag{32}$$

For x slightly bigger than $x_{k\omega}$ (we will simply write $x>x_{k\omega}$) both $a(x,k\omega)$ and $a(x,j\omega)$ vanish. In the interval

$$x_{k\omega} \geqslant x > x_{j\omega} \tag{33}$$

only $a(x,j\omega)$ vanishes and the response is of the form

$$\mathrm{Re}(e^{i\kappa}c_k a(x,k\omega)e^{ik\omega t}). \tag{34}$$

On this whole interval, the frequency $k\omega/2\pi$ can be heard.

At this point we should be specific about hearing a frequency. There are two different ways in which the cochlea perceives frequencies. First by analyzing the frequency content of the afferent neurons (firing frequencies). This is independent of the place x at which the analysis is made. The second way is to identify the place x with its resonating frequency ω_x. If the neurons connecting to the place x are active (nonrandom firing) and the neurons connecting to neighboring places show less activity (their firing shows less regularity) then the cochlea hears the frequency ω_x. This is a criterion which can be used at low sound levels. For higher intensities, it is just the firing rate of the neuron, which should be maximal for the neuron which connects to the place x.

In our example, at all places x in the interval $x_{k\omega}\geqslant x>x_{j\omega}$ the firing frequencies of the neurons let the cochlea hear the frequency $k\omega/2\pi$.

In order to see what happens at $x=x_{j\omega}$ (and for all x in some interval below $x_{j\omega}$) we make the assumption that $a(x,\omega)$ is dilation invariant:

$$a(x,\omega) = a(x-\gamma\log c, c\omega). \tag{35}$$

So far we only assumed this to be true for the phase $\vartheta(x,\omega)$ of the basilar membrane filter.

That this remains true for the amplification factor $a(x,\omega)$—at least within some limitations—will be discussed elsewhere.

For $x\leqslant x_{j\omega}$ both frequencies $k\omega/2\pi$ and $j\omega/2\pi$ are present. Moreover the phase relation of the two harmonics remains the same for all $x\leqslant x_{j\omega}$, at least as long as $a(x,k\omega)$ and $a(x,j\omega)$ are of a non-negligible size [i.e., the approximation $\vartheta(x,\omega)=\kappa-\mu\omega/\omega_x$ holds].

Let us see what happens at the place $x_{j\omega}$ (cf. Fig. 1).

Since

$$x_{j\omega} = x^* - \gamma\log j\omega = x^* - \gamma\log\omega - \gamma\log j = x_\omega - \gamma\log j, \tag{36}$$

$$a(x_{j\omega},k\omega) = a(x_\omega - \gamma\log j + \gamma\log k,\omega), \tag{37}$$

the amplification factors become

$$a(x_{j\omega},j\omega) = a(x_\omega,\omega) := a_{jj}, \tag{38}$$

$$a(x_{j\omega},k\omega) = a\left(x_\omega - \gamma\log\frac{j}{k},\omega\right) := a_{jk}. \tag{39}$$

Clearly $a_{jk}<a_{jj}$, since $a_{jj}=\max_{x,\omega}a(x,\omega)$.

If the signal is chosen such that

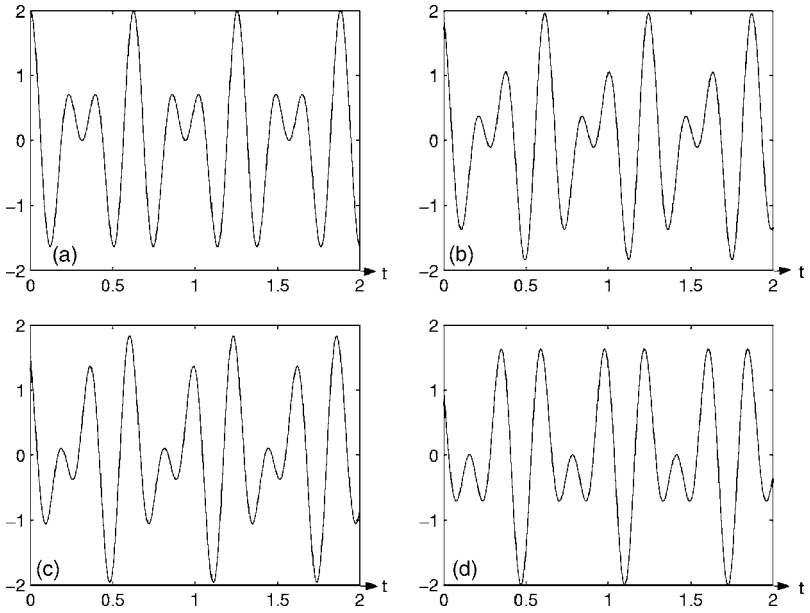


FIG. 2. Two harmonics with frequency ratio 2:3, varying relative phase. See Eq. (42). The values for φ in (a), (b), (c), (d) are $0, \pi/6, \pi/3, \pi/2$.

$$|c_k|a_{jk} = |c_j|a_{jj} := A \tag{40}$$

then $u(x_{j\omega}, t+\Delta t)$ is the superposition of two harmonics, both of the same amplitude A,

$$\begin{aligned} u(x_{j\omega}, t+\Delta t) &= \mathrm{Re}(e^{i\kappa}\{c_k a_{jk} e^{ik\omega t} + c_j a_{jj} e^{ij\omega t}\}) \\ &= \mathrm{Re}(e^{i\kappa} A\{e^{i\varphi_k} e^{ik\omega t} + e^{i\varphi_j} - e^{ij\omega t}\}), \end{aligned} \tag{41}$$

$$h(t) = \mathrm{Re}(e^{i2\omega t+\varphi} + e^{i3\omega t}). \tag{42}$$

If the phases

$$\varphi_k = \arg c_k, \tag{43}$$

$$\varphi_j = \arg c_j \tag{44}$$

are varied the oscillation pattern for $u(x_{j\omega}, t+\Delta t)$ changes (see Fig. 2). In particular, if the phases are chosen such that

$$\varphi_k + \kappa \equiv \varphi_j + \kappa \equiv 0 \bmod 2\pi, \tag{45}$$

then for $t = n2\pi/\omega$, $n \in \mathbb{Z}$, the two harmonics simultaneously reach their maximal value A. The firing histogram for the nerve fibers with best frequency $j\omega$ will therefore contain the period $2\pi/\omega$ and the contribution of this period will be essential. This means the cochlea hears the basic frequency $\omega/2\pi$ at the place $x_{j\omega}$.

All this remains true for x in an interval below $x_{j\omega}$, since the condition

$$\varphi_k + \kappa \equiv \varphi_j + \kappa \equiv 0 \bmod 2\pi \tag{46}$$

is independent of x. Note, however, that the equivalent of

$$|c_k|a_{jk} = |c_j|a_{jj} \tag{47}$$

will hold only approximately for those nearby values of x. But this will suffice for the present argumentation.

Other choices of the phases φ_k and φ_j can be made such that

$$\max_t \mathrm{Re}\{A e^{i\kappa+i\varphi_k} e^{ik\omega t} + A e^{i\kappa+i\varphi_j} e^{ij\omega t}\} \tag{48}$$

will remain substantially below $2A$ (see Fig. 2). The two harmonics will not reach their maximal value A simultaneously. In such a situation the period $2\pi/\omega$ will hardly be present in the firing histogram.

A listener who is confronted with these two different situations will hear the basic frequency $2\pi/\omega$ with the first choice of the phases but not with the second.

2. Example 2: Mistuned consonances

In music it is well known that two tones with a frequency ratio slightly different from $k:j$, k and j being small integers, give rise to a beat sensation. With our present approach it seems clear that the beat rate will have to agree with the number of wave form variations per second. In the historical survey by Plomp (1967) this idea is attributed to Smith with reference to the book "Harmonics" 1749.

For completeness we include here a short calculation for the beat frequency. The result is due to De Morgan (1864).

Consider a linear combination of two sinusoids of equal amplitude with frequency ratio close to $k:j$ ($k<j$, k and j no common divisor)

$$f(t) = \mathrm{Re}\{c_k e^{i(k\omega+\delta)t} + c_j e^{i(j\omega+\delta)t}\}. \tag{49}$$

The distortion $\delta/2\pi$ is chosen to be small, typically 1–3 Hz.

The frequency change produces a change of the individual periods:

$$\frac{2\pi}{k\omega} - \frac{2\pi}{k\omega+\delta} = \frac{2\pi\delta}{k\omega(k\omega+\delta)}, \tag{50}$$

$$\frac{2\pi}{j\omega} - \frac{2\pi}{j\omega+\delta} = \frac{2\pi\delta}{j\omega(j\omega+\delta)}. \tag{51}$$

Within a rough common period $2\pi/\omega$ the differences add up to

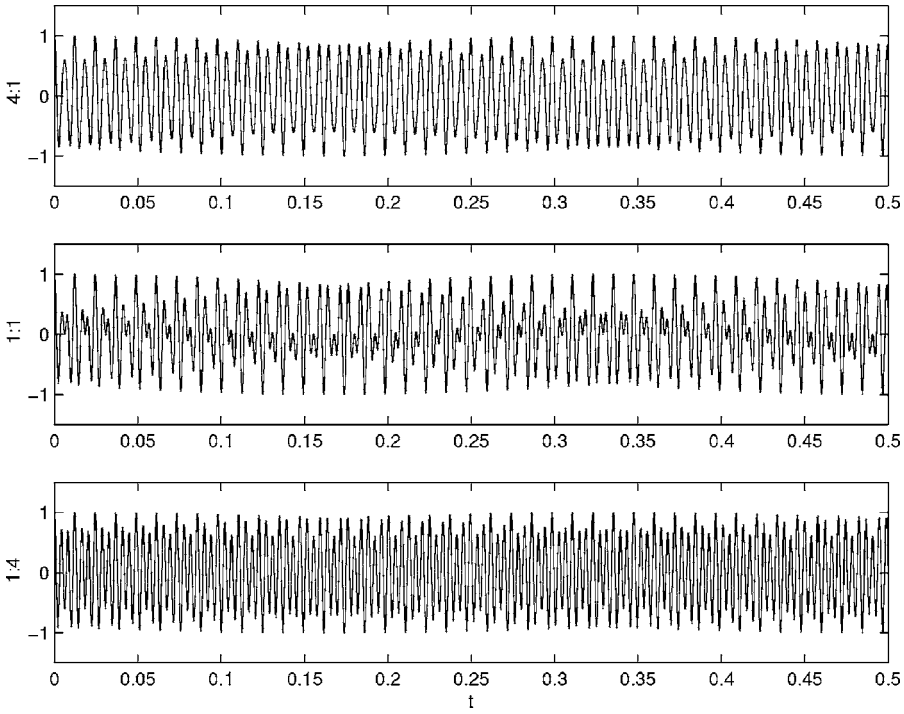


FIG. 3. Oscillation pattern for harmonics with frequency ratio close to 2:3. The various amplitude ratios are indicated. Relative scale.

$$k\frac{2\pi\delta}{k\omega(k\omega+\delta)} - j\frac{2\pi\delta}{j\omega(j\omega+\delta)} \cong \frac{(j-k)2\pi\delta}{(k\omega+\delta)(j\omega+\delta)}. \quad (52)$$

The difference of the undisturbed periods is

$$\frac{2\pi}{k\omega} - \frac{2\pi}{j\omega} = \frac{2\pi}{\omega}\frac{j-k}{jk} \quad (53)$$

and the number of periods $2\pi/\omega$ required to make up for this difference is

$$l = \frac{2\pi}{\omega}\frac{j-k}{jk} / \frac{(j-k)2\pi\delta}{(k\omega+\delta)(j\omega+\delta)} = \frac{1}{\delta\omega jk}(k\omega+\delta)(j\omega+\delta)$$

$$\cong \frac{\omega}{\delta}. \quad (54)$$

The period for the beat is therefore

$$l \cdot \frac{2\pi}{\omega} = \frac{2\pi}{\delta} \quad (55)$$

and the beat frequency $\delta/2\pi$ equals the distortion frequency. It is remarkable that in this approximation the beat frequency depends neither on ω nor on the numbers k and j.

Within a beat period the vibration pattern undergoes a cycle. If as in example 1 the constants c_k and c_j of the mistuned consonances are chosen such that

$$|c_k|a_{jk} = |c_j|a_{jj} = A \quad (1 < k < j) \quad (56)$$

then the vibration of the basilar membrane at $x_{j\omega}$ is of the form

$$u(x_{j\omega}, t+\Delta t) = \mathrm{Re}\{Ae^{i\kappa}(e^{i\varphi_k + i(k\omega+\delta)t} + e^{i\varphi_j + i(j\omega+\delta)t})\}. \quad (57)$$

Within a beat period the pattern changes in such a way that the dominance of the basic frequency $\omega/2\pi$ goes on and off (see Fig. 3). In fact on a whole interval below $x_{j\omega}$ the same changes of the pattern lead to the same effect. This explains the sensation of a beat. Observe that for natural sounds with basic frequencies $(k\omega+\delta)/2\pi$ and $(j\omega+\delta)/2\pi$ this phenomena is repeated for all combinations of the higher harmonics.

3. Example 3: The sound of a violin

On a violin the sound is produced by bowing across the strings. Under such a movement, the vibration of the strings takes a sawtooth shape. Therefore the Fourier spectrum of the sound generated essentially contains all the higher harmonics, with amplitudes decreasing and with all harmonics in phase. It is here conjectured that for violins of a high quality these elements also show in the projected sound.

Assume then that the violin sound is of the form

$$f(t) = \mathrm{Re}\sum_k c_k e^{ik\omega t} \quad (58)$$

with

$$|c_1| > |c_2| > |c_j| > \cdots \quad (59)$$

and

$$\arg c_k = \mathrm{const}. \quad (60)$$

For simplicity we will assume $\kappa=0$ and we will take $\arg c_k=0$ for all k. For each combination of higher harmonics we find ourselves in the situation of example 1. Consequently, the basic frequency $\omega/2\pi$ is present along the cochlea at all places $x \leqslant x_\omega$.

If the orchestra plays with a solo violinist then large parts of the cochlea are covered by the orchestral sound. However the fact that the sound of the solo violin extends over the whole range $x \leqslant x_\omega$ of the cochlea and this with coherent phase is probably what allows the ear to recognize the sound of the solo violin.

II. WAVELETS

A. Amplification

The basis for the considerations in the first part was the invariance of the phase function

$$\vartheta(x,\omega) = \vartheta(x - \gamma \log c, c\omega). \tag{61}$$

This was a direct consequence of Zweig's postulate that ϑ only depends on ω/ω_x.

A similar statement seems to hold for the amplifying factor $a(x,\omega)$. Here however, this factor depends on the sound level of the incoming signal. The following considerations therefore only apply within a narrow band of sound intensity.

At a fixed intensity of sound we look for an invariance property which allows for more flexibility. As such we require that a satisfies

$$a(x - \gamma \log c, c\omega) = \chi(c) a(x,\omega) \tag{62}$$

with χ a function of the dilation parameter. With such an ansatz we still respect the experimental evidence, that $a(x,\omega)$ and $a(x,c\omega)$ qualitatively look the same apart from the shift by $\gamma \log c$ along the x axis but that $a(x,c\omega)$ tends to be smaller than $a(x,\omega)$ if $c>1$.

Combining phase and amplification factor we take the dilation invariance of the basilar membrane filter in the form

$$\hat{g}(x - \gamma \log c, c\omega) = \chi(c) \hat{g}(x,\omega) \tag{63}$$

as our starting point.

We immediately obtain the information that χ has to be a character of the dilation group. Applying the invariance property successively with c_1 and c_2 and comparing the result with the equation for the product $c=c_1c_2$ gives

$$\begin{aligned}\hat{g}(x - \gamma \log(c_1c_2), c_1c_2\omega) &= \chi(c_1)\chi(c_2)\hat{g}(x,\omega) \\ &= \chi(c_1c_2)\hat{g}(x,\omega)\end{aligned} \tag{64}$$

and consequently

$$\chi(c_1c_2) = \chi(c_1)\chi(c_2). \tag{65}$$

This is the defining property for characters. For the dilation group, the characters are of the form

$$\chi(c) = c^\sigma \tag{66}$$

with a fixed $\sigma \in \mathbb{C}$. We therefore must have

$$\hat{g}(x - \gamma \log c, c\omega) = c^\sigma \hat{g}(x,\omega). \tag{67}$$

But then the function

$$h(x,\omega) := \omega_x^{-\sigma} \hat{g}(x,\omega) = e^{-\sigma(x^* - x)/\gamma} \hat{g}(x,\omega) \tag{68}$$

satisfies the original invariance law:

$$\begin{aligned}h(x - \gamma \log c, c\omega) &= \exp\{-\sigma(x^* - x + \gamma \log c)/\gamma\} \hat{g}(x \\ &\quad - \gamma \log c, c\omega) \\ &= c^{-\sigma} \exp\{-\sigma(x^* - x)/\gamma\} c^\sigma \hat{g}(x,\omega) \\ &= h(x,\omega).\end{aligned} \tag{69}$$

Observe that if we take $\sigma \in \mathbb{R}$ then nothing changes in the invariance behavior of the phase $\vartheta(x,\omega)$. However if σ is complex valued, this influences the invariance behavior of the phase.

B. The basilar membrane filter as a wavelet transform

In her book "Ten Lectures on Wavelets" Daubechies (1992) observes that the dilation invariance of the basilar membrane filter shows that $u(x,t)$ is expressed as a wavelet transform.

We will in the following assume that the phase $\vartheta(x,\omega)$ of $\hat{g}(x,\omega)$ is a function of ω/ω_x only (Zweig), and that the amplifying factor $a(x,\omega)$ also has this property (cf. the preceding section) when multiplied by the factor $\exp\{-\sigma(x^* - x)/\gamma\}$, $\sigma \in \mathbb{R}$,

$$h(x,\omega) = e^{-\sigma(x^* - x)/\gamma} a(x,\omega), \tag{70}$$

$$h(x - \gamma \log c, c\omega) = h(x,\omega). \tag{71}$$

This invariance relation is equivalent to the statement that h is a function of ω/ω_x only.

Our assumption therefore is that $\hat{g}(x,\omega)\exp\{-\sigma(x^* - x)/\gamma\}$ is a function of ω/ω_x.

There thus exists a complex valued function $\hat{\psi}$ such that

$$\hat{\psi}\left(\frac{\omega}{\omega_x}\right) = e^{-\sigma(x^* - x)/\gamma} \overline{\hat{g}}(x,\omega), \quad \omega > 0. \tag{72}$$

In the resulting formula

$$u(x,t) = e^{-\sigma(x^* - x)/\gamma} \mathrm{Re} \sqrt{\frac{2}{\pi}} \int_0^\infty \overline{\hat{\psi}}\left(\frac{\omega}{\omega_x}\right) \hat{f}(\omega) e^{i\omega t} d\omega \tag{73}$$

we use the parameter transformation

$$c = \frac{1}{\omega_x} = e^{-(x^* - x)/\gamma}, \tag{74}$$

$$x = x^* + \gamma \log c. \tag{75}$$

This leads to

$$u(x^* + \gamma \log c, t) = \frac{c^{-\sigma}}{\sqrt{c}} \mathrm{Re} \sqrt{\frac{2}{\pi}} \int_0^\infty \sqrt{c}\, \overline{\hat{\psi}}(c\omega) \hat{f}(\omega) e^{i\omega t} d\omega. \tag{76}$$

Extending the function $\hat{\psi}$ to negative values in a Hermitian symmetric way

$$\hat{\psi}(\omega) = \overline{\hat{\psi}}(-\omega), \quad \omega < 0 \tag{77}$$

the integral can be written as an integral over $\mathbb{R}$,

$$u(x^* + \gamma \log c, t) = \frac{c^{-\sigma}}{\sqrt{c}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \sqrt{c}\, \overline{\hat{\psi}}(c\omega) \hat{f}(\omega) e^{i\omega t} d\omega. \tag{78}$$

The function $\overline{\hat{\psi}} \in L_{\mathrm{sym}}$ is the transform of a real valued function ψ and the inverse Fourier transform of $\sqrt{c}\,\overline{\hat{\psi}}(c\omega)$ is the function $(1/\sqrt{c})\psi(-t/c)$. The convolution theorem

$$(h * \hat{f}) = \sqrt{2\pi}\hat{h} \cdot \hat{f} \tag{79}$$

applies with $\hat{h}(\omega) = \sqrt{c}\,\overline{\hat{\psi}}(c\omega)$:

$$u(x^* + \gamma \log c, t) = \frac{c^{-\sigma}}{\sqrt{c}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{c}} \psi\left(\frac{s-t}{c}\right) f(s)\,ds. \tag{80}$$

By definition, this last integral is the wavelet transform $W_\psi f$ of f with wavelet ψ.

With the notation

$$\psi_{c,t}(s) = \frac{1}{\sqrt{c}} \psi\left(\frac{s-t}{c}\right) \tag{81}$$

this can be written as

$$W_\psi f(c,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{c}} \psi\left(\frac{s-t}{c}\right) f(s)\,ds = \int f(s)\,\psi_{c,t}(s)\,ds$$
$$= (f, \psi_{c,t}). \tag{82}$$

The cochlear process is aptly described by a wavelet transform

$$u(x^* + \gamma \log c, t) = \frac{c^{-\sigma}}{\sqrt{2\pi c}} W_\psi f(c,t), \quad c > 0 \tag{83}$$

with wavelet ψ.

Linearity and invariance with respect to the affine group are the essential ingredients of this deduction. Both the cochlear process and the wavelet transform go with the affine group.

Our concern is the wavelet ψ. In the case $\sigma = 0$ the Fourier transform $\hat{\psi}$ is given as

$$\hat{\psi}(\omega/\omega_x) = \overline{\hat{g}}(x,\omega) = e^{-i\,\vartheta(x,\omega)} a(x,\omega). \tag{84}$$

For the phase $\vartheta(x,\omega)$ we use the expressions from Sec. I:

$$\vartheta(x,\omega) = \kappa - \mu \frac{\omega}{\omega_x}. \tag{85}$$

The amplification factor $a(x,\omega)$ can be described by a function of a single variable:

$$a(x,\omega) := b(\omega/\omega_x). \tag{86}$$

For the wavelet $\hat{\psi}$ this gives

$$\hat{\psi}(\omega/\omega_x) = e^{-i\kappa + i\mu\omega/\omega_x} b(\omega/\omega_x) \tag{87}$$

for positive values of α, whereas the Hermitian symmetric extension of $\hat{\psi}$ to all values is

$$\hat{\psi}(\omega/\omega_x) = e^{-i\kappa\ \mathrm{sign}(\omega) + i\mu\omega/\omega_x} b(\omega/\omega_x). \tag{88}$$

In the case $\sigma \in \mathbb{R}$ we have

$$\hat{\psi}(\omega/\omega_x) = e^{-\sigma(x^*-x)/\gamma} e^{-i\,\vartheta(x,\omega)} a(x,\omega) \tag{89}$$

and we can now set

$$e^{-\sigma(x^*-x)/\gamma} a(x,\omega) = b_\sigma(\omega/\omega_x). \tag{90}$$

This gives the wavelet

$$\hat{\psi}(\omega/\omega_x) = e^{-i\kappa\ \mathrm{sign}(\omega) + i\mu\omega/\omega_x} b_\sigma(\omega/\omega_x). \tag{91}$$

Note that the additional factor $c^{-\sigma}$ appears in the wavelet transform.

ACKNOWLEDGMENT

My thanks go to Roy Patterson and all the members of his laboratory for their hospitality during my visit in Cambridge and for their lively interest in this work.

Daubechies, I. (**1992**). *Ten Lectures on Wavelets* (SIAM, Philadelphia).
De Morgan, A. (**1864**). "On the beats of imperfect consonances," Trans. Cambridge Philos. Soc. **2**, 129–141.
Geisler, D. (**1998**). *From Sound to Synapse* (Oxford University Press, New York).
Greenwood, D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Liberman, C. E. (**1982**). "The cochlear frequency map for the cat: Labeling auditory nerve-fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.
Lighthill, J. (**1981**). "Energy flow in the cochlea," J. Fluid Mech. **106**, 149–213.
Plomp, R. (**1967**). "Beats of mistuned consonances," J. Acoust. Soc. Am. **42**, 462–474.
Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.
Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 317–332.
Zweig, G. (**1976**). "Basilar membrane motion," *Cold Spring Harbor Symposia on Quantitative Biology*, pp. 619–633.

Model for vocalization by a bird with distensible vocal cavity and open beak

Neville H. Fletcher[a)]
Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

Tobias Riede and Roderick A. Suthers
School of Medicine, Jordan Hall, Indiana University, Bloomington, Indiana 47405

(Received 4 August 2005; revised 23 November 2005; accepted 29 November 2005)

Some birds make use of a distensible oral cavity to produce nearly pure-tone song. Songbirds such as the Northern cardinal (*Cardinalis cardinalis*) have a muscularly distended oropharyngeal-esophageal cavity between the top of the trachea and the open beak. The present paper analyzes the acoustics of this vocal system. It is shown that the resonance of the oropharyngeal-esophageal cavity, vented through the beak, introduces a dominant peak in the radiation efficiency, the frequency of which can be adjusted by varying the volume of the cavity, the beak gape, and perhaps the position of the tongue in the mouth. To produce nearly pure-tone song, the bird adjusts the frequency of this peak to coincide with the fundamental of the syringeal oscillation. The present paper provides the acoustical analysis underlying this behavior. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159434]

PACS number(s): 43.64.Tk, 43.80.Ka, 43.70.Aj [JAS] Pages: 1005–1011

I. PURE-TONE BIRDSONG

While much birdsong consists of complex sounds with harmonic partials generated by air flow through the vibrating syringeal valve and a formant envelope imposed by resonances of the upper vocal tract, some birds are able to produce almost pure-tone songs. There has been speculation in the past that these songs might be produced by a completely different aerodynamic whistle mechanism, but experiments have now demonstrated that, at least in many cases, the syrinx is vibrating normally, though perhaps remaining partially open through its whole vibration cycle instead of closing for a part of it (Larsen and Goller, 1999). While this nonclosing valve oscillation certainly reduces the content of higher harmonics in the syringeal flow, some sort of narrow band-pass filter mechanism is required in the upper vocal tract to produce the single-tone calls with greatly attenuated upper harmonic content actually observed.

In the case of Ring doves (*Streptopelia risorea*), the mechanism involves an inflated esophagus and a closed beak. The upper esophagus is inflated to rather large volume so that its walls are very thin. The air flow from the syrinx is lightly filtered by the first resonance of the trachea, which may be adjusted by partial closure of the glottis, and then led into the inflated esophagus, the beak remaining closed during vocalization (Beckers *et al.*, 2003; Riede *et al.*, 2004). The air enclosed in the inflated esophagus, coupled to the vibrational motion of the sac walls, constitutes the necessary narrow-band filter, and sound is radiated directly by the wall vibrations (Fletcher *et al.*, 2004). The analysis indicates that the filter frequency is not very sensitive to esophagus inflation provided it is moderately inflated, so that its tuning is not significantly altered during the rather brief duration of the "coo" song. This means, however, that the bird is not able to change its song pitch by an appreciable amount, and it remains close to the inflated sac resonance at about 600 Hz in the Ring doves studied.

Songbirds, however, such as the Northern cardinal (*Cardinalis cardinalis*), possess an actively distended oropharyngeal esophageal cavity (hereafter referred to as the OEC) in the vocal tract, do not close their beaks during song, and are able to produce pure tones that can be swept rapidly over a wide range of frequency, even from 1 to 9 kHz, as described by Riede *et al.* (2006) in a companion experimental paper and illustrated in the penultimate figure of the present paper. For these birds, most of the OEC is embedded in tissue so that its walls do not vibrate much, and the song is then essentially that radiated through the beak. The present paper analyzes the acoustic performance of such a vocal system, examines the extent to which any OEC resonance is dominant over tracheal resonances, and explores the degree to which the OEC resonance can be modified by such variations in the anatomical parameters as could be made by the bird during song.

II. MODEL DESCRIPTION

The aim of this model is to present the basic features of the vocal tract of the bird in such a way that a calculation of its acoustic performance is possible without the intervention of too many unknown anatomical parameters. The dimensions of the trachea and other anatomical features are known from x-ray images of spontaneously singing birds and from anatomical dissections. The only adjustable parameters are taken to be the (a) constriction of the glottis, (b) the volume of the OEC, and (c) the acoustic loading produced by the

[a)]Author to whom correspondence should be addressed. Electronic mail: neville.fletcher@anu.edu.au

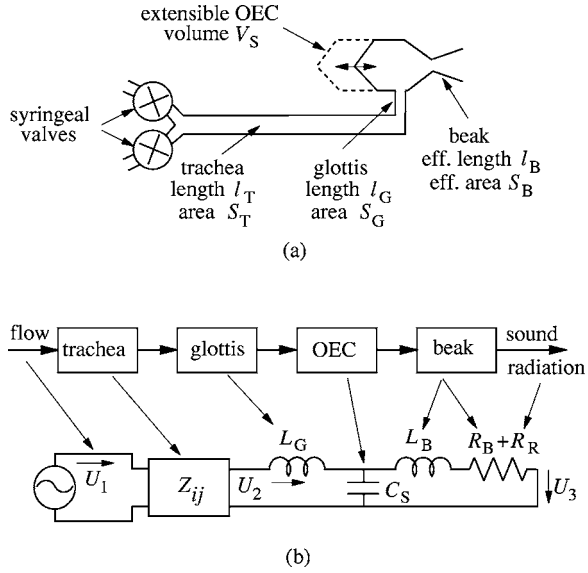


FIG. 1. (a) Sketch of the vocal system of the bird that is studied. Symbols define the dimensions used in the model. (b) Elements of the acoustic system describing the bird vocal system in (a), together with the electrical network analog circuit used in the calculation.

mouth and beak, reduced to a single parameter related to beak gape. Since in the cardinal the syrinx is very close to the tracheal bifurcation, resonances of the bronchi below the bifurcation, which were considered in an earlier birdsong model (Fletcher and Tarnopolsky, 1999), can be neglected. It is then assumed that the active syringeal valve behaves as a high-impedance acoustic source injecting a constant volume flow amplitude at frequency $f=\omega/2\pi$ into the base of the trachea.

The physical model for the vocal system is shown in Fig. 1(a), the acoustic model in the upper part of Fig. 1(b), and the electric analog network computational model in the lower part of Fig. 1(b). The meanings of the symbols are defined in the diagram and will be discussed in what follows. An electric network analog model is used since this is common in acoustics and makes the calculation simple. In this model, acoustic pressure is represented by electric potential and acoustic volume flow by electric current (Fletcher, 1992). The electrical analogs of short constrictions then appear as inductors (L) and cavities that are a good deal smaller than the wavelength as capacitors (C), while tubes are four-terminal impedance elements Z_{ij}.

A much more detailed model of the vocal anatomy could easily be formulated, recognizing in particular that the OEC is a tubelike structure with the mouth at one end and the glottis entering about half-way along one side. Such a model might also take into account the detailed taper of the OEC at the mouth end. To do all this, however, would add a great deal more complication to the analysis and require several extra parameters and would make little difference to the calculated results for frequencies less than that at which the OEC is half a wavelength long. This complication becomes significant at frequencies above about 6 kHz for a fully distended OEC and above about 12 kHz when it is collapsed. In view of the approximations made in other aspects of the analysis, this complication has not been included, but the frequency limitations of the simple model are shown when the results are displayed.

Using the symbols as defined in Fig. 1, the network equations can be simply solved to show that the acoustic flow U_3 through the beak at frequency ω for an input flow U_1 is given by

$$U_3 = \frac{-jZ_{12}\omega C_S U_1}{1-(1-\omega^2 L_B C_S + j\omega R_B C_S)(1-\omega^2 L_G C_S + j\omega C_S Z_{22})}, \tag{1}$$

where $j=\sqrt{-1}$ as usual, and the impedances still have to be defined in terms of the physical dimensions of the system. The necessary results are as follows, defined in each case in terms of the dimensions shown in Fig. 1(a) (Fletcher, 1992):

$$Z_{11} = Z_{22} = -jZ_0 \cot kl_T, \tag{2}$$

$$Z_{21} = Z_{12} = -jZ_0 \operatorname{cosec} kl_T, \tag{3}$$

where $Z_0=\rho c/S_T$ with ρ the density of air and c the speed of sound in air, and $k=\omega/c-j\alpha$ where the attenuation coefficient α is given by $\alpha\approx 10^{-5}\beta\omega^{1/2}a^{-1}$ with a being the effective radius of the trachea, given by $S_T=\pi a^2$, and β is an extra loss coefficient ($1<\beta<10$) for the walls of the trachea (Fletcher, 1992). When the complex nature of k is included specifically, it follows that

$$Z_{11} = Z_{22} = -Z_0 \frac{\sinh(\alpha l_T)\cosh(\alpha l_T) + j\sin(\omega l_T/c)\cos(\omega l_T/c)}{[\sin(\omega l_T/c)\cosh(\alpha l_T)]^2 + [\cos(\omega l_T/c)\sinh(\alpha l_T)]^2}, \tag{4}$$

$$Z_{21} = Z_{12} = -Z_0 \frac{\cos(\omega l_T/c)\sinh(\alpha l_T) + j\sin(\omega l_T/c)\cosh(\alpha l_T)}{[\sin(\omega l_T/c)\cosh(\alpha l_T)]^2 + [\cos(\omega l_T/c)\sinh(\alpha l_T)]^2}. \tag{5}$$

The glottal constriction, assumed to have effective length l_G and area S_G, has an analog inductance $L_G=\rho l_G/S_G$. It should be noted that, since the length of the distended OEC is not always small compared with the sound wavelength, a really careful model would include the possibility of internal resonances within the OEC itself. These are difficult to model, since the OEC does not have a simple shape, but a first-order model in which it is treated as a simple cylinder with the glottis near the mid-point is outlined in the Appendix. Calculations using this model show that

there is, as expected, a sharp minimum in the output from the beak when the OEC length is exactly half the sound wavelength. Above this frequency there is an additional series of maxima when the OEC length is an odd number of quarter wavelengths of the sound. This model is not pursued in detail, since it is not anatomically accurate, but its findings indicate that the results of the present calculation are not accurate at high frequencies for a greatly extended OEC and, for this reason, the calculated results to be shown later are all truncated at the frequency above which they become unreliable.

Possible vibration of the walls of the OEC contributes an effective inductance $L_W \approx \rho_W T_W / A_W^2$, where ρ_W is the mass density of the tissue surrounding the cavity, T_W is the thickness of this tissue, and A_W is the total area of the cavity walls. An important anatomical distinction between the cardinal and the dove is that the OEC is embedded in other tissue, giving a large value of T_W, while the expanded esophagus of the dove is essentially exposed when inflated. This wall inductance L_W is in parallel with the impedance of the air passage through the beak. However, its magnitude is so much greater than that of the beak vent that it can validly be neglected. Any wall vibration will also contribute an additional loss to the system but, in the absence of any data, this is neglected.

The beak is much more complicated, since its impedance depends upon the nature of the constriction at its junction with the OEC, the position of the tongue and the extent to which this constricts the passage through the mouth into the beak, and the dimensions and opening angle of the beak itself. It is well known that the mouth geometry is crucial in determining the frequencies of the formant resonances in human speech, and the same has been found to be true of parrots (Patterson and Pepperberg, 1994; Beckers *et al.*, 2004). As discussed in detail elsewhere (Fletcher and Tarnopolsky, 1999) and verified by experiment for a simplified beak geometry, the beak contributes an inertive (i.e., inductive) load to the OEC opening. For a typical conical beak about five times as long as its base diameter, the result is that the beak behaves as an open tube equal to its base in diameter and with an effective length l_{eff} that varies between about half its geometrical length when it is nearly closed down to about one-tenth of this length when its tip is opened to more than about one-third of the beak length. This acoustic length also depends upon frequency f, and a regression to fit the experimental results has the form

$$l_{\text{eff}} \approx 0.05 l_B + 0.03 \frac{l_B^2}{g\lambda} \approx 0.05 l_B + 10^{-4} \frac{f l_B^3}{g}, \tag{6}$$

where g is the tip opening and λ is the sound wavelength. The second form of writing assumes that dimensions are given in meters. This expression, which is reasonably accurate for $g < l_B$, that is for a gape angle of less than about 60°, and for a frequency f such that $f < 500/l_B$, is illustrated in Fig. 2. [Note that the expression (6) has been modified slightly from that in the original paper by insertion of an extra factor $10 l_B$ to scale the frequency properly with beak length.] If the root diameter of the beak is d_B, then its acoustic inertance is about

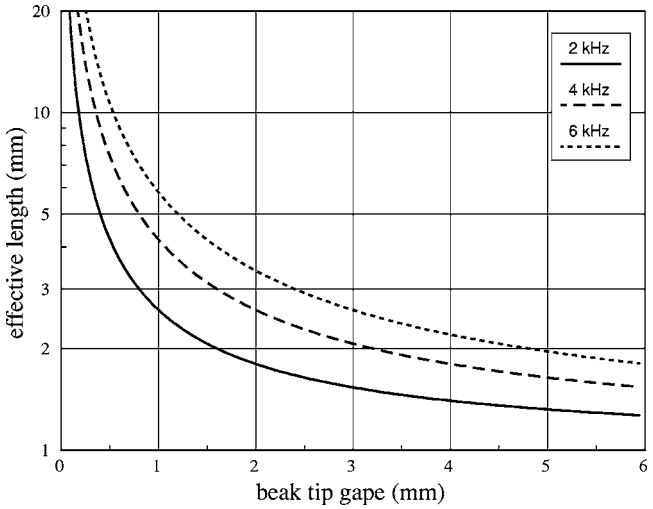


FIG. 2. Plot of the relation (6) giving the effective acoustic length as a function of frequency f and tip gape g, for the case of a beak 20 mm in length and 4 mm in effective diameter at its base.

$$L_B \approx \frac{4 l_{\text{eff}}}{\pi d_B^2}. \tag{7}$$

It can be seen from Fig. 2 that the effective acoustic length of the beak is very small unless it is almost closed and saturates to a value of a little more than 1 mm for a widely open beak, this being largely due to the radiation impedance. This figure is not strictly applicable to the beak of the cardinal, since here the root diameter is more like one-fifth of the beak length, and the actual tip gape during singing ranges from nearly 0 to about 4 mm, but the general behavior will be very similar. This leads to the conclusion that beak opening is not the controlling factor influencing the dominant frequency of the song syllable, but rather a combination of the OEC volume and the beak gape.

There may be an additional inertive load contributed by the constriction caused by the tongue, depending upon how this is raised during song, which may well be correlated with the beak gape. For simplicity, however, this will be simply included with the beak inertance as defined by its effective length and root diameter. This is accomplished by defining a parameter G, which will in what follows be called the gape parameter, through the relation

$$L_B = \frac{\rho l_B}{S_B} = \frac{\rho}{G}. \tag{8}$$

This parameter G is then measured in meters (or more conveniently in millimeters) and increases as the beak is opened or the tongue lowered so as to reduce its constricting effect. Reasonable anatomical considerations suggest that, for the birds considered, G lies in the range 1–7 mm. The value of G is positively correlated with, but not numerically equal to, the physical opening of the beak tip.

This gape parameter G will be used in the illustrative calculations in the next section. Note that, even if the mouth and beak were so widely open that each could be assumed to have zero length for a trial calculation, G could still not exceed about 10 mm because of the inertive component of

the radiation impedance of a simple opening, which is equivalent to that of a tube matching this opening and about 0.3 times its diameter in length (Fletcher, 1992). Most of the acoustic tuning effect of the beak and mouth thus occur when the beak is nearly closed, in agreement with the findings of Nelson *et al.* (2004). From the curves of Fig. 2, the parameter G, while positively correlated with actual beak gape, does not vary linearly with this quantity, most of the variation in G occurring when the beak is nearly closed. In addition, from (8) the variation in G could be largely caused by variation in the root area S_B of the beak or by variation in the shape of the OEC where it joins the beak, which is included in this parameter. Care should therefore be taken when interpreting the parameter G in the results to be displayed later.

The acoustic impedance of the mouth and beak, as calculated above, is simply the inertive component, and to this must be added a resistance $R_B \ll Z_0$, where $Z_0 = \rho c / S_T$ is the characteristic impedance of the trachea. This resistance is due largely to viscous losses in the various constrictions around the tongue and the sharp edges of the beak, since the radiation resistance is negligibly small in comparison. Quantification of this resistance is difficult since it depends upon things such as the sharpness of the edges of the beak and the shape and roughness of the tongue. It must therefore be regarded as a rather poorly specified adjustable parameter, but its value simply affects the sharpness of some of the resonance peaks. Finally, there is the radiation resistance R_R, as shown in Fig. 1. Since $R_R \ll R_B$, it can however be neglected for the purposes of the network analysis.

With these definitions, the analysis is complete, though the expressions for the impedance coefficients Z_{ij} must be expanded to take account of the complex nature of k as indicated in (4) and (5), and Eq. (1) must then be simplified to give the magnitude $|U_3|$ of the acoustic flow from the beak. Since the equivalent cross-section diameter of the beak aperture is generally small compared with the sound wavelength, the acoustic radiation resistance R_R varies approximately as ω^2 so that the radiated acoustic power at frequency ω is proportional to $\omega^2 U_3^2$ (Fletcher, 1992). Expressed as usual in decibels relative to some arbitrary reference, the sound level is given by $20 \log_{10}(\omega U_3) + \text{const}$. In the calculations to follow it will be assumed that the magnitude of U_1 is independent of frequency so that a spectral response function can be calculated. For an actual bird call, the syringeal mechanics will determine the relative amplitudes of the harmonics of the fundamental in the syringeal flow (Fletcher, 1988) and, from this and the response function derived above, the spectrum of the radiated song can be calculated.

III. RESULTS AND CONCLUSIONS

It is straightforward to implement a computer program to calculate the relative radiated power level $20 \log_{10}(\omega U_3)$ as a function of frequency for particular values of the parameters that the bird varies during song, namely the OEC volume V_S, the beak gape parameter G, and perhaps the glottal constriction S_G, and for reasonable assumed values for the other parameters. For the particular case studied here, the Northern cardinal, these parameters are shown in Table I.

TABLE I. Assumed parameter values.

length of trachea	l_T	48 mm
diameter of trachea	d_T	1.6 mm
loss factor of trachea	β	3
length of glottis	L_G	2 mm
diameter of glottis	d_G	1 mm
length of OEC	l_S	5–20 mm
volume of OEC	V_S	0.3–2 ml
loss resistance of beak	R_B	10^5 Pa s m^{-3}

Figures 3 and 4 show the calculated acoustic power output for a constant glottal flow at a given frequency. As is most clearly shown in Fig. 3, there are four acoustic resonant reinforcement peaks. The frequencies of the peaks at about 1.2, 4, and 7 kHz, indicated by arrows, are associated with tracheal resonances and vary rather little with changes in the gape parameter G, while the frequency of the most promi-

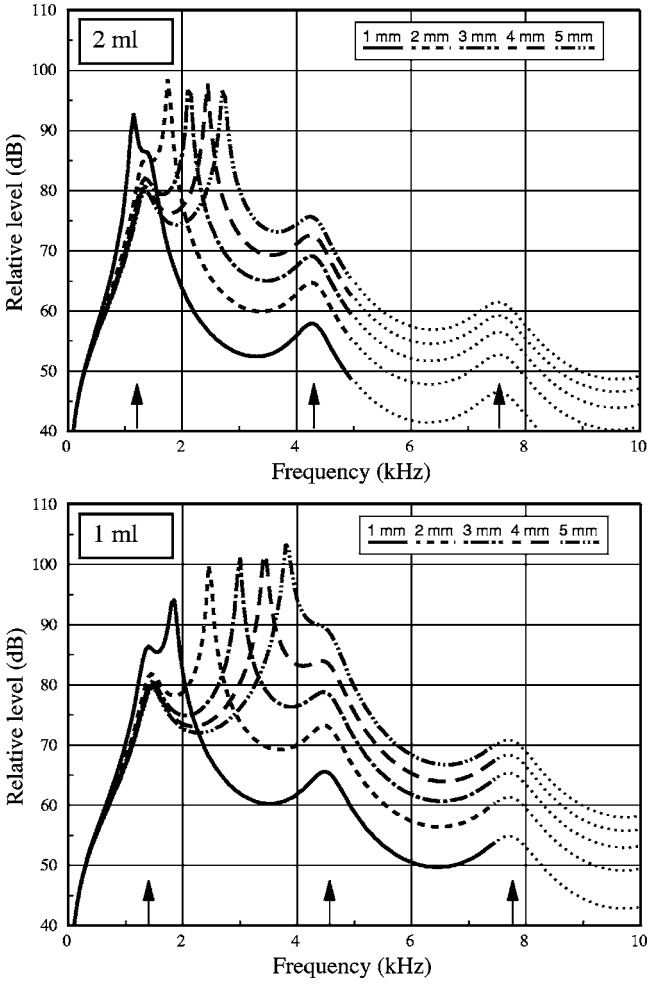


FIG. 3. Relative radiated power level as a function of frequency, for a constant input volume flow at the syrinx, with a OEC volume of 2 or 1 ml and various values of the gape parameter G. Other parameters are as shown in Table I. The location of the variable OEC resonance peak is shown to depend upon both beak gape and OEC volume, while the frequencies of the three tracheal resonances, near 1.5, 4.5, and 7.5 kHz and indicated by arrows, are little affected by these parameters unless there is a near coincidence with the OEC peak. The predictions of the model are not accurate above frequencies for which the wavelength is less than about twice the OEC length, and the curves are shown dotted in this region.

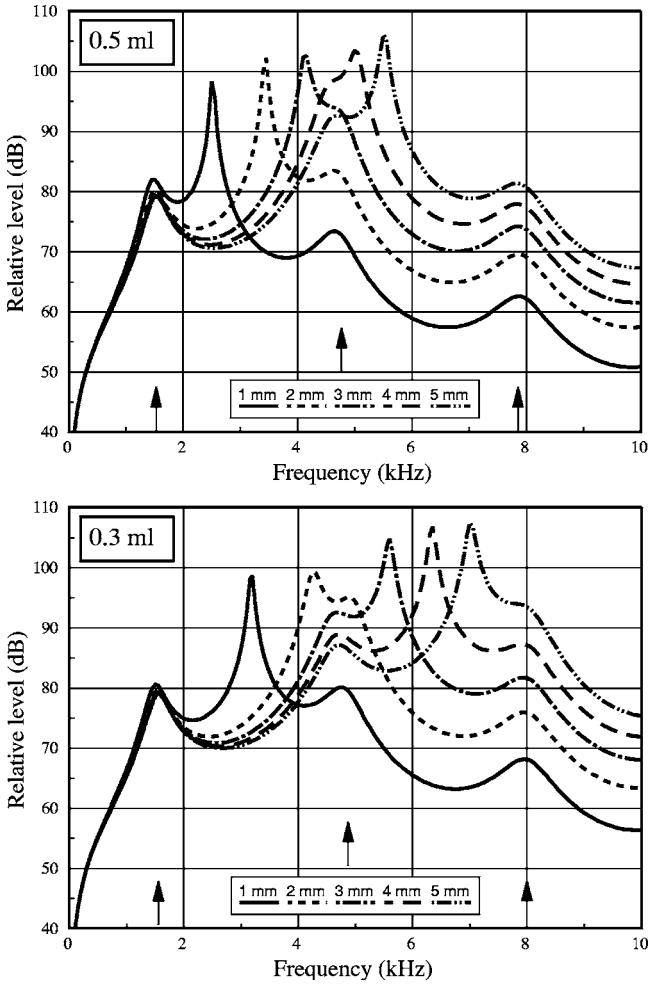


FIG. 4. Plots similar to those of Fig. 3 for OEC volumes of 0.5 or 0.3 ml as functions of beak gape parameter G. Other comments are as for Fig. 3.

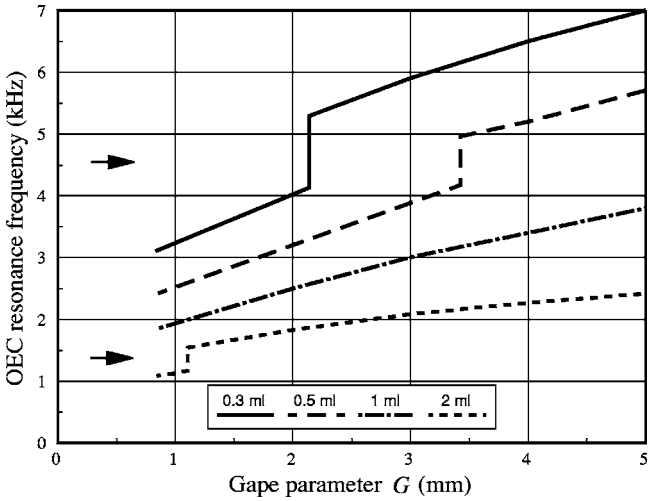


FIG. 5. Frequency of the dominant resonance peak of the vocal tract as a function of beak gape parameter G for four different OEC volumes. The tracheal resonances are at about 1.5 and 4.5 kHz, as indicated by the arrows. The curves do not reach the 7.5-kHz resonance.

nent resonance in the range 1–4 kHz, in contrast, depends strongly on G, as well as on the OEC volume V_S. Figure 4 shows the results of similar calculations with two smaller assumed OEC volumes, allowing the dominant resonance to be moved as high as 7 kHz. Note that, when the frequency of this variable resonance approaches that of one of the tracheal resonance peaks, as shown most clearly in Fig. 4(a) for $G=3$ mm and in Fig. 4(b) for $G=2$ mm, there is an acoustic interaction and, over a small parameter range, the heights of the two peaks are nearly equal before the dominant resonance frequency jumps from one to the other.

The relative prominence of the tracheal peaks depends upon the wall losses in the trachea, the value assumed in the calculation being that for a moderately smooth tube of uniform diameter. In addition to this, however, since the OEC acts to some extent as a low-pass filter, its presence reduces the prominence of all tracheal resonances above the frequency of the OEC resonance. This effect is greatest when the OEC volume is large, as is readily seen by comparing Figs. 3 and 4. The prominence of the variable OEC resonance peak depends mostly upon the assumed value of the resistive losses in the mouth and beak. Using the assumed parameter values, however, the dominant resonance peak is always that associated with the OEC. The frequency of this peak as a function of the gape parameter G is shown in Fig. 5, for the four OEC volumes used in the calculation. The sudden jumps in frequency occur when the dominant resonance passes through the frequency of one of the tracheal resonances. It is clear that, by varying the mouth configuration, beak opening, and OEC volume over the reasonable ranges assumed, the bird can tune the prominent vocal tract resonance over a wide range from below 1 kHz to about 7 kHz. At some frequencies there may be two or more combinations of OEC volume and beak gape that provide the necessary resonance frequency. For example, referring to Fig. 5, an OEC resonance and thus a song frequency match at 3.5 kHz could be achieved with an OEC volume of 0.3 ml and a beak gape parameter of 1.5 mm, with an OEC volume of 0.5 ml and a gape of 2.5 mm, or an OEC volume of 1 ml and a gape of 4.5 mm. The bird is able to choose whichever of these options, or others in between, is most convenient for the syllable being sung.

The bird will presumably match the vibration frequency of its active syringeal valve with that of this variable resonance, thus ensuring a large vocal output at this fundamental frequency, which can be varied over at least the range 1–7 kHz. It is reasonable that the two syringeal valves may differ slightly in thickness and musculature so that one is used for the lower and one for the higher notes.

Because the air flow through the syringeal valve is nonlinear, this generates all harmonics of the fundamental frequency in the flow U_1, as has been discussed elsewhere (Fletcher, 1988), the harmonic amplitudes generally decreasing smoothly in amplitude with increasing frequency at about 6 dB/oct (i.e., as $=\omega^{-1}$). On top of this, the properties of the vocal OEC lead to a deemphasis of frequencies above the OEC resonance, as is particularly noticeable in Fig. 3. Since there are no higher resonances of the OEC to consider, and since it is unlikely that any of the subsidiary tracheal resonances will happen to coincide with upper harmonics of the excitation frequency, these upper harmonics of the syringeal flow will be very weak in the radiated spectrum, and the song will have a predominantly pure-tone nature. Exami-

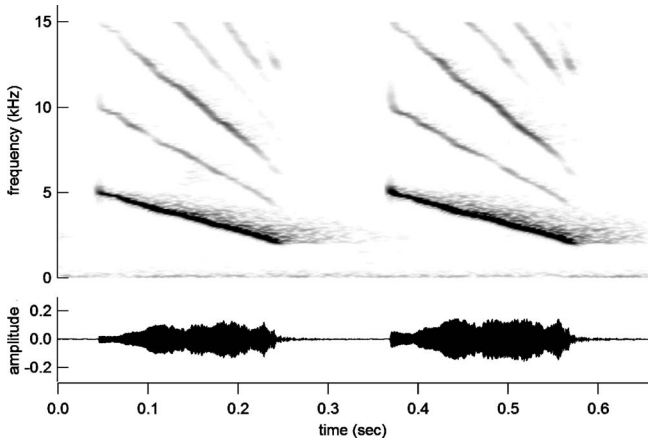


FIG. 6. Waveform and time-resolved spectrum of two syllables sung by a cardinal. The dominant fundamental is associated with the adjustable resonance of the OEC, while the two broad formant bands at about 7, 10, and 14 kHz appear to be associated with the fixed tracheal resonances revealed by the analysis.

nation of Figs. 3 and 4 shows that the peak associated with the OEC resonance is 10–15 dB higher than the nearest tracheal resonance peak. This implies a corresponding advantage in the power radiated at this frequency and explains why evolution may have favored this development.

Additional supporting data for the model are shown in Fig. 6, which gives the waveforms and time-resolved spectra for two downward sweeping syllables of the cardinal's song. There is clearly one dominant frequency component of the sweep, decreasing in frequency from about 6 to 3 kHz, which we identify with the major resonance of the OEC, as discussed above. In addition, however, there are emphasized formant bands in the much weaker higher harmonics of this fundamental, centered round 7, 10, and 14 kHz, which might reasonably be taken to be associated with the fixed tracheal resonances identified in the analysis. The lack of quantitative agreement between these frequencies and those calculated is not very significant, since the model is primarily concerned with the major OEC resonance and its accuracy declines at high frequencies.

One apparent problem of a behavioral nature is the necessity for the bird to adjust the frequency of its syringeal vibration to match that of the major OEC resonance. While it is true that immature birds do find difficulty in producing substantially pure-tone songs, perhaps for this very reason, it does appear to be a rather difficult task. Fortunately there is a simple solution to this apparent dilemma. In all the discussion above it has been assumed that the bird produces a complex harmonic acoustic flow through the syringeal valve and that this signal is then simply filtered by the vocal tract, a model known as the "source-filter approximation." While this is a good approximation in the case of birds such as the raven (Fletcher, 1988), since the frequency of the fundamental is much less than that of the first vocal tract resonance, this is no longer true for the birds under consideration here. Instead, when the frequency of a major vocal tract resonance is close to that of the vibrating valve source, there is a major feedback interaction that tends to shift the frequency of the valve oscillation into close match with that of the resonance

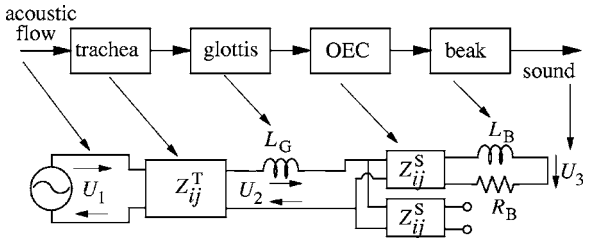


FIG. 7. The acoustic system showing the OEC as a cylinder with the glottal entry at its mid-point, together with the electrical network used to model the behavior of this system.

(Fletcher, 1978). The reason for this is that the valve oscillation is driven by the pressure difference across it, and this is strongly influenced on the downstream side by vocal tract resonances. What this implies, therefore, is that the major tuning action is that of distension of the OEC coupled to beak gape, and that, provided the vibrational frequency of the syringeal valve is adjusted to approximate agreement, the coupling mechanism will lock its frequency to that of the resonance. In this connection, "approximate agreement" means to within about $\pm f_S/Q$ where f_S is the OEC resonance frequency and Q is the quality factor of the undriven valve, which is probably less than 10 because of losses when the valve closes. The required frequency match accuracy is probably thus not as strict as ±10%.

It should perhaps be emphasized once again that the model presented here has used the simplest anatomical approximations possible in order to present the fundamentals of the operation of the vocal system of birds such as the cardinal. Many refinements could be introduced, at the expense of introducing additional parameters, but should not affect the overall predictions in the frequency ranges shown.

APPENDIX: EXTENDED TREATMENT

The physical model for the vocal system shown in Fig. 1(a) can be refined by treating the OEC as an extended object with the trachea entering at the glottis half way along its length. The acoustic model for this system is shown in the upper part of Fig. 7, and the electric analog network computational model in the lower part of Fig. 7. Because the OEC, when fully distended, is nearly 30 mm in length, it is necessary to represent it by a pair of four-terminal impedances with the flow from the glottis entering between them. One of these elements leads to the beak, while the other is terminated by a closed wall with only a very small hole leading to the non-distended part of the esophagus, so that there is essentially no acoustic flow at its remote end, which is therefore shown as an open circuit.

Using the symbols as defined in Fig. 1 but with Z_{ij}^T representing the impedance coefficients for the trachea and Z_{ij}^S, those for the half-length of the OEC, the network equations can be solved to show that the acoustic flow U_3 through the beak at frequency ω for an input flow U_1 is given by

$$U_3 = \frac{Z_{11}^S Z_{12}^S Z_{12}^T}{[Z_{11}^S + 2Z_{11}^T + 2Z_G][Z_{11}^S Z_B + (Z_{11}^S)^2] - (Z_{12}^S)^2(Z_{11}^S + Z_{11}^R + Z_G)}, \quad \text{(A1)}$$

where the impedances still have to be defined in terms of the physical dimensions of the system.

This system cannot be modeled until the physical dimensions of the OEC have been fully specified, and this is difficult since they vary smoothly with OEC distention. As a representative case, the OEC can be treated as a simple cylinder so that the coefficients Z_{ij}^S have the same form as those given in (4) and (5) for the trachea, with appropriate dimensional changes. Implementation of the model then shows that it gives results that are essentially identical with those of the simple cavity model discussed in the main text for frequencies such that the total OEC length is less than about 0.4 times the sound wavelength. Above this frequency there are additional maxima and minima in the response. This model is not pursued in detail, since it is not anatomically realistic. A more realistic model could, however, be constructed using the result (A1) and values of the impedance coefficients Z_{ij}^S appropriate to the actual OEC geometry if this were known in detail. The actual calculation of these coefficients would, however, be quite complicated.

Beckers, G. J. L., Nelson, B. S., and Suthers, R. A. (**2004**). "Vocal-tract filtering by lingual articulation in a parrot," Curr. Biol. **14**, 1592–1597.

Beckers, G. J. L., Suthers, R. A., and ten Cate, C. (**2003**). "Pure-tone birdsong by resonance filtering of harmonic overtones," Proc. Natl. Acad. Sci. U.S.A. **100**, 7372–7376.

Fletcher, N. H. (**1978**). "Mode locking in nonlinearly excited inharmonic musical oscillators," J. Acoust. Soc. Am. **64**, 1566–1569.

Fletcher, N. H. (**1988**). "Bird song—a quantitative acoustic model," J. Theor. Biol. **135**, 455–481.

Fletcher, N. H. (**1992**). *Acoustic Systems in Biology* (Oxford U. P., New York).

Fletcher, N. H., and Tarnopolsky, A. (**1999**). "Acoustics of the avian vocal tract," J. Acoust. Soc. Am. **105**, 35–49.

Fletcher, N. H., Riede, T., Beckers, G. J. L., and Suthers, R. A. (**2004**). "Vocal tract filtering and the 'coo' of doves," J. Acoust. Soc. Am. **116**, 3750–3756.

Larsen, O. N., and Goller, F. (**1999**). "Role of syringeal vibrations in bird vocalization," Proc. R. Soc. London, Ser. B **266**, 1609–1615.

Nelson, B. S., Beckers, G. J. L., and Suthers, R. A. (**2004**). "Vocal tract filtering and sound radiation in a songbird," J. Exp. Biol. **208**, 297–308.

Patterson, D. K., and Pepperberg, I. M. (**1994**). "A comparative study of human and parrot phonation: Acoustic and articulatory correlates of vowels," J. Acoust. Soc. Am. **96**, 634–648.

Riede, T., Beckers, G. J. L., Blevins, W., and Suthers, R. A. (**2004**). "Inflation of the esophagus and vocal tract filtering in ring doves," J. Exp. Biol. **207**, 4025–4036.

Riede, T., Suthers, R. A., Fletcher, N. H., and Blevins, W. E. (2006). "Songbirds tune their vocal tract to the fundamental frequency of their song," Proc. Natl. Acad. Sci. U.S.A. (submitted).

Induced loudness reduction as a function of exposure time and signal frequency

Eva Wagner
Institute for Hearing, Speech & Language and Department of Speech-Language Pathology & Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

Bertram Scharf[a)]
Department of Psychology and Institute of Hearing, Speech & Language, Northeastern University, Boston, Massachusetts 02115 and INSERM EPI 9902, Laboratoire d'Otologie et de NeuroOtologie, Université d'Aix-Marseille II, Marseille, France

(Received 12 November 2004; revised 27 November 2005; accepted 28 November 2005)

Induced loudness reduction (ILR) is the decline in the loudness of a weaker tone induced by a preceding stronger tone. In this study we investigate how ILR depends on exposure time and signal frequency. For 12 listeners, successive magnitude estimation was used to measure the loudness of 70-dB-SPL test tones, presented with and without preceding 80-dB-SPL inducer tones at the same frequency. Experiment 1 measured the evolution of ILR over time at 0.5 kHz. The results suggest that ILR may begin after a single inducer presentation, and increases over at least 2 to 3 min as the inducer and test tones are repeated every few seconds. Following the cessation of the inducer, the recovery of loudness is slow and still incomplete after 1 min. Experiment 2 extended the measurements to additional signal frequencies. The results show that the amount of ILR and its evolution over time are approximately the same at frequencies from 0.5 to 8 kHz. Similarly, loudness matching showed no effect of frequency on ILR, which averaged 8.2 dB. These findings, together with previously noted similarities among ILR, ipsilaterally induced loudness adaptation, and temporary loudness shift, indicate that loudness reduction induced by stronger sounds is a very common phenomenon. [DOI: 10.1121/1.2159430]

PACS number(s): 43.66.Cb [AJO] Pages: 1012–1020

I. INTRODUCTION

Stronger tones usually cause weaker tones that follow in time to decrease in loudness. This induced loudness reduction or ILR (e.g., Scharf *et al.*, 2002) is a ubiquitous phenomenon whenever certain constraints on the relation between the stronger inducer tone and the weaker test tone are met. Relative to the test tone, the inducer must be 5 to 20 or 30 dB more intense (Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003), must not differ too much in frequency (Marks and Warner, 1991; Marks, 1994; also in a related paradigm, Charron and Botte, 1988), must not be too brief relative to the test duration (Nieder *et al.*, 2003), and must end at least several hundred milliseconds before the onset of the test tone (Scharf *et al.*, 2002; Arieh and Marks, 2003a). Under these optimal conditions, ILR averages as much as 10 or 11 dB. Our knowledge about these relative stimulus conditions is much better than about absolute conditions. However, it does seem that ILR is strongest for inducers between 60 and 90 dB SPL, decreasing quickly at lower levels and remaining stable at higher levels (Scharf *et al.*, 2005). As to signal duration, essentially nothing is known about durations longer than 1 s, except in related paradigms (see Sec. IV). With respect to signal frequency, some results (e.g., Marks and Warner, 1991; Mapes-Riordan and Yost, 1999) suggest that ILR may be stronger at lower than at higher frequencies, but this remains most uncertain. Finally, although the dependence of ILR on the temporal interval between the inducer and the test tone has been mapped out by Arieh and Marks (2003a) and recovery has been tracked by Arieh *et al.* (2005), the development of ILR over time is unknown. We do not know how many times and/or for how long the inducer tone (and test tone) must be presented before ILR begins and then reaches the asymptote. In this paper we focus on these questions concerning the effect of signal frequency on ILR and the development of and recovery from ILR.

A word is in order about the nature of ILR. Early measurements, mainly by magnitude estimation, of what was then called the slippery context effect (e.g., Marks and Warner, 1991) and later loudness recalibration (Marks, 1994) reflected the thinking that ILR required a fairly complex set of stimuli; it seemed that at least two different frequencies had to be presented over unequal intensity ranges. However, a test tone at a single frequency presented first at a higher and then at a lower level is all it takes to demonstrate ILR (Scharf *et al.*, 2003). Measurements of ILR based on loudness matching do employ a remote, second frequency as a comparison tone, but the second frequency plays no role in ILR (Marks, 1993, Experiment 15; Parker and Schneider, 1994; Scharf *et al.*, 2002).

Although the mechanism underlying ILR is unknown, it almost certainly does not result from response biases, procedural quirks, or perceptual distortions—ILR is a true sensory

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

effect. Most telling of all in this respect is an experiment by Arieh and Marks (2003b), which omits any kind of loudness judgment. Given that listeners respond more rapidly to louder sounds than to softer sounds, Arieh and Marks measured the reaction time under conditions that normally lead to ILR. Both simple and choice reaction times were slower (and were associated with more errors) to tones that underwent ILR than to tones not so softened. Added to this is the report by Scharf *et al.* (2003) that listeners show similar amounts of ILR under two very different methods; the correlation between ILR measured by loudness matching (a 2AFC adaptive procedure) and by magnitude estimation was 0.67.

It is to be noted that even if it is a sensory effect, ILR differs markedly among listeners but varies little within listeners. For example, Scharf *et al.* (2003) reported a correlation of 0.79 between the ILR in the left ear and that in the right ear. Equally high correlations held between ILR measured on the same listeners for test-tone levels of 65 and 75 dB SPL and between ipsilateral ILR (inducer and test in same ear) and contralateral ILR (inducer and test in different ears). In other words, whether measured in different ears, at different levels, or under different input conditions, the amount of ILR varies little for a given listener. As to the dependence of ILR on temporal properties, only the delay between the inducer and test tone has been thoroughly studied. Arieh and Marks (2003a) showed that ILR requires that the test tone follow the inducer by at least 200 ms; beyond 200 ms, the amount of ILR increases as the interstimulus interval lengthens to about 600 ms and appears to remain stable thereafter, up to intervals of 3 s, perhaps even up to intervals of 10 s (Mapes-Riordan and Yost, 1998). These measurements were based on many repetitions of the inducer and test tones in a matching procedure. Hence, we do not know how quickly ILR develops after the onset of the first inducer. Nor do we know whether the evolution and amount of ILR is different at different signal frequencies. This paper provides answers to these questions.

II. EXPERIMENT 1: DEVELOPMENT AND DECLINE OF ILR FOR A 0.5 KHZ TONE

Experiment 1 measured ILR as a function of the length of the exposure to the inducer and test tones. Such measurements are most readily carried out by the method of successive magnitude estimation (ME), which provides an excellent tool for determining how sensory magnitude changes over time. This procedure has been used successfully in numerous studies of loudness adaptation (e.g., Scharf *et al.*, 1983; Miskiewicz *et al.*, 1993) and loudness decruitment (e.g., Canévet and Scharf, 1990; Schlauch *et al.*, 1995; see Sec. IV) as well as in analogous studies in vision (Teghtsoonian *et al.*, 2000) and taste (Meiselman, 1968). In successive ME, observers estimate at brief intervals of time, e.g. every 5 s, the perceived magnitude of a continuous, physically unchanging stimulus (in studies of adaptation) or of a stimulus decreasing monotonically in intensity (in studies of decruitment). In the application of successive ME to ILR (cf. Scharf *et al.*, 2003), listeners judge the loudness of a test tone that is repeated over and over again, preceded on every trial by an inducer tone. The stimulus paradigm is like that in many recent studies that used a matching procedure to measure ILR (e.g., Mapes-Riordan and Yost, 1999; Nieder *et al.*, 2003; Arieh and Marks, 2003a); the main difference is that for matching purposes, a comparison tone at a distant frequency follows the test tone on every trial, whereas under ME, tones at only a single frequency are presented. Under matching, ILR is equal to the level of the comparison tone when judged equal in loudness to a test tone presented alone (in a series of baseline trials) minus the level of the comparison tone when the test tone is preceded on every trial by an inducer. This single measure, based on a series of trials, reveals little or nothing about the development of ILR over time. Mapes-Riordan and Yost (1999) did give one example of how the loudness matches evolved from trial to trial, but the judgment on each trial depends heavily on the level of the comparison tone on that particular trial and always entails a dichotomous judgment, louder or not louder. In ME, the listener estimates the loudness of the test tone on each trial so that the amount of change in loudness, if any, is directly measured. The estimations on trials when an inducer accompanies the test tone are compared to prior estimations when the test tone had been presented alone. A decrease in the estimation upon the introduction of the inducer provides a direct measure of ILR. Biases inherent in ME are, for the most part, overcome by averaging among a sufficient number of listeners.

A. Method

1. Listeners

Twelve listeners, six females, participated in this experiment. None had a history of hearing difficulties, and all had pure-tone thresholds at or below 15 dB HL (ANSI, 1996) at octave frequencies between 0.25 and 8 kHz. They ranged in age from 19 to 32 years. Five listeners were paid for their services; the rest were members of the laboratory, including the first author (L8).

2. Stimuli

The test tones and inducers always had a frequency of 0.5 kHz and lasted 200 ms, including raised-cosine rise and fall times of 6.67 ms. The level of the inducer was 80 dB SPL; the level of the test tone was 70 dB SPL.

3. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded the listeners' responses, and executed the procedure. The sample rate was 48 kHz. The output of the 16-bit D/A converter (TDT DD1) was attenuated (TDT PA4), low-pass filtered (TDT FT5, f_c = 20 kHz, 135 dB/oct), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of a Sony MDR-V6 headset. Listeners sat in a sound-attenuating booth (Acoustic Systems), and the stimuli were presented monaurally to the preferred ear. For routine calibration, the output of the headphone amplifier was led to a 16-bit A/D converter (TDT DD1) such that the computer

TABLE I. Loudness estimations by each of the 12 listeners as a function of the time (and trial) at which the test tone was presented. (The inducer was present from trial 3 through trial 43. Bold numbers represent trials with no inducer. Last line gives for each listener the ratio of the estimation on trial 2 to the estimation on trial 43.)

Trial	Time (s)	Inducer	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	Geomean
2	**8**	**No**	**5**	**10**	**8**	**4**	**3**	**5**	**5**	**10**	**4**	**2**	**4**	**1**	**4.3**
3	12	Yes	4	1	7	4	4	5.1	3	10	4	2	4	1.6	3.5
8	32	Yes	3	3	4	3	4	4.9	2	9	3	2	3	1.4	3.2
13	52	Yes	2	3	5	3	3	4.5	2	9	3	1.5	3	1.5	3.0
18	72	Yes	3	3	5	3	3	4.3	2	8	3	2	2	2	3.1
23	92	Yes	2	2	5	2	4	4.8	2	7.5	3	2	2	3	2.9
28	112	Yes	3	2	4	2	3	5.5	2	7	3	2	3	2	2.9
33	132	Yes	2	2	5	2	3	5.8	2	6	3	2	2	2	2.8
38	152	Yes	2	1	3	2	3	5.2	2	6	3.5	2	2	4	2.7
43	172	Yes	2	2	4	1	2	5	1.5	6	4	2	2	4.2	2.6
48	**192**	**No**	**4**	**1**	**4**	**2**	**3**	**5**	**1**	**6**	**3**	**2**	**3**	**2**	**2.6**
53	**212**	**No**	**3**	**2**	**3**	**2**	**3**	**5.5**	**1**	**6**	**3.5**	**2**	**4**	**1**	**2.6**
58	**232**	**No**	**3**	**2**	**4**	**2**	**3**	**5.6**	**1**	**6**	**4**	**3**	**4**	**1.1**	**2.8**
63	**252**	**No**	**3**	**2**	**4**	**2**	**4**	**6**	**1**	**6**	**4**	**3**	**4**	**0.9**	**2.9**
Trial 2/Trial 43			2.5	5	2	4	1.5	1	3.3	1.7	1	1	2	0.2	1.65

could sample the waveform, calculate its spectrum and rms voltage, and display the results before each block of trials. The SPLs reported below assume a frequency-independent output at the earphone of 116 dB SPL for an input of 1 V rms.

4. Procedure

Listeners estimated loudness by assigning numbers to represent the loudness magnitude. They were told that they could use all positive numbers, including decimals, which they typed into a small terminal. They were free to select any number for the first estimation as long as they felt that the number magnitude matched the loudness. Emphasis was put on matching a number to the loudness on every trial. Each estimation was called for by the word "judgment" on the terminal screen.

On each of the first two trials of a series of 63 trials, listeners heard a single 70-dB-SPL tone burst. They estimated the loudness of the tone burst on the second trial. On the third trial and on each of the next 40 trials, an 80 dB inducer preceded the 70 dB test tone by 1.8 s. Listeners estimated the loudness of the weaker test tone; they were to ignore the louder inducer. They continued to judge the 70 dB test tone on every fifth trial for a total of 9 estimations with an inducer present. Beginning with the 44th trial, the inducer was no longer presented and was replaced by 200 ms of silence. Listeners estimated the loudness of the now unaccompanied test tone four more times, on every fifth trial until the series ended on the 63rd trial. (The first three columns of Table I give the experimental sequence.) The magnitude estimations were completed in less than 10 min. Each listener judged the series of tones once only. Owing to the persistence of ILR, a second series could not be run on the same day. In any case, repeating the series even on a different day would have served little purpose. Owing to the small range over which loudness changed and the similarity of the sounds, listeners would be likely to remember their earlier judgments and repeat them. As noted above, listeners are consistent in their loudness judgments under ILR conditions. Finally, we were interested in average results as in most previous studies of sequential effects on loudness that used successive magnitude estimation and that tested each listener once under each condition (e.g., Hellman *et al.*, 1997).

B. Results and discussion

Table I presents the 14 loudness estimations by each of the 12 listeners as a function of time. The first estimation and the last four, printed in bold font, were made on trials with no inducer. Our principal measure of ILR is the ratio of the estimation at trial 2 with no inducer to that at trial 43, the last trial with an inducer. The geometric mean of these individual ratios is 1.65 (which is the same as the ratio of the corresponding mean estimations in the last column of Table I). Thus, on average, loudness was reduced by 40% at the end of 41 repetitions of the inducer and test tone, which took approximately 160 s. Taking the mean ratio of 1.65 and assuming the standard exponent of 0.6 for the function that relates the loudness of the 0.5 kHz tone to sound pressure, we calculate an equivalent ILR of 7.2 dB. This value is less than the 11 dB reported by Nieder *et al.* (2003) and Mapes-Riordan and Yost (1999) for an 80 dB inducer and 70 dB test tone also at 0.5 kHz, but measured by loudness matches in a two-alternative forced-choice (2AFC) procedure. We suggest in the discussion of experiment 2 that this discrepancy comes about probably because our listeners constricted the range of numbers that they assigned to represent loudness. Experiment 2 provides a way to take this constriction into account.

Figure 1 plots the geometric means as a function of the time from the first trial. The standard error, as a percentage of the geometric mean, varied over time from 13% to 21%. The initial estimation of the 70 dB tone with no inducer-averaged 4.3. A few seconds later, the very first estimation with a preceding 80 dB inducer was down to 3.5. The estimations continued to fall over the next minute, and then more slowly thereafter until the estimation of the loudness after the last inducer at 3 min was 2.6. Recovery during the

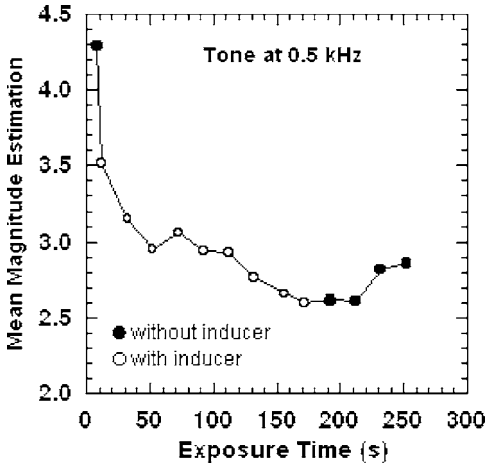


FIG. 1. Geometric means of 12 individual magnitude estimations of the loudness of 70-dB-SPL tones with (open symbols) and without (closed symbols) a preceding 80-dB-SPL inducer as a function of exposure time from the first trial.

final minute without an inducer was slow and incomplete with the final estimation equal to only 2.9, notably less than the initial estimation of 4.3.

Overall, these data suggest that for some listeners ILR at 0.5 kHz occurs already after a single presentation of the inducer, with loudness declining rapidly over the first minute as the inducer and test tones are repeated every few seconds. (The effect is clearer in experiment 2, where nearly half the listeners reported a decline in loudness after the very first inducer, with few reporting an increase.) Although the amount of ILR increases little over the next 2 to 3 min of inducer trials, the loudness remains depressed. Removal of the inducer on trial 44 does not lead to an immediate increase in loudness; rather, loudness recovers slowly and is still well below its initial preinducer value even after more than 1 min with no inducer.

With respect to the development of ILR, our results are in reasonable agreement with previous reports. Marks (1993, experiment 14) inferred a rapid development of ILR (or "differential context effect," as he then called it) from magnitude estimations of 0.5 and 2.5 kHz tones. He found that tones were judged relatively louder as soon as they were switched from a series of mostly stronger tones into a series of mostly weaker tones; the same tones were judged relatively softer after the converse shift. However, as in the matching data of Mapes-Riordan and Yost (1999), the inference of the speed with which ILR begins was very approximate. As to recovery, Arieh *et al.* (2005) found that little more than 2 min were required for complete recovery from exposure to 50 ms, 80 dB inducers repeated 40 times at intervals of 1 s. The results of our experiment (see also experiment 2) suggest that more than 2 min is required for full recovery to the initial loudness. The slower recovery may arise from differences in stimulus conditions: our tones were 200 ms compared to their 50 ms; the exposure to our 43 inducers was stretched over 160 s whereas the exposure to the 40 inducers of Arieh *et al.* lasted 40 s, and our test-tone level was 70 dB compared to their 60 dB level.

We note that Arieh, one of the reviewers of this paper, suggested that we make sure that listeners would not give progressively smaller numbers to a repeated tone burst, even in the absence of the inducer. Such a control could show that listeners do not simply tend to give smaller and smaller numbers when presented a succession of tones. From our experience with successive magnitude estimation, we are confident that, on average, listeners would give approximately the same number to a series of repeated tones in the absence of intervening stronger tones. For example, the second author has shown, by this same method of successive magnitude estimation, that at most signal levels, listeners repeatedly assign the same number over many minutes to the loudness of a constant-intensity sound (cf. Scharf, 1983). Perhaps most telling is the fact that the listeners in the present study began to give progressively *larger* numbers to a series of repeated tone bursts, shortly after the inducer ceased. Nonetheless, we do believe that the control suggested by Professor Arieh is well worth doing.

Although our own results show clearly how ILR progresses over exposure time, we do not know whether the progression is the same for inducer tones presented at other rates that would yield fewer or more exposures over a given time. The question of how many exposures to an inducer are required to reach maximum or near maximum ILR is relevant to assessing its possible role in various psychophysical procedures in which stronger and weaker tones are presented together for varying amounts of times and in various numbers. In a procedure such as adjustment in which the listener usually sets the variable sound to higher levels on several interspersed trials, do those exposures cause the variable to sound softer on succeeding trials? Could ILR then be part or even all of the reason why, a comparison sound is generally set to a higher level to match a standard when it is varied in level than when it is fixed? [For a striking example of this discrepancy, see Zwicker (1958).] These and similar questions await more data on just how ILR depends on the number and frequency of exposure to inducers.

III. EXPERIMENT 2: ILR AT SIX SIGNAL FREQUENCIES

Previous measurements of ILR provide little information about the effect of signal frequency. Measurements by magnitude estimation have been restricted mainly to tones at 0.5 and 2.5 kHz (e.g., Arieh and Marks, 2003b). Marks and Warner (1991) and Marks (1994) did test at several frequencies, but their main experimental variable was the frequency difference between inducer and test tone, which they did not seek to dissociate from absolute frequency. Most measurements of ILR by loudness matching have been at a single frequency, 0.5 kHz (e.g., Nieder *et al.*, 2003) or 2.5 kHz (e.g., Arieh and Marks, 2003a). Only Mapes-Riordan and Yost (1999) reported loudness matches at more than one frequency. They measured more ILR at 0.5 kHz than at 2.5 kHz. Additional information about the effect of frequency can be gleaned from measurements of ILR in the guise of loudness enhancement (see Scharf *et al.*, 2002), measurements made at 1 kHz (e.g., Plack, 1996) and 5 kHz (e.g., Elmasian and Galambos, 1975). Comparisons across all these studies do not suggest any consistent dependency of

the amount of ILR on signal frequency. However, these various studies do not permit a definitive conclusion because they differed too much in procedure, and because only one study made measurements at more than a single frequency on the same listeners. Accordingly, we applied the method of successive magnitude estimation to tones at six different frequencies. At each frequency, we measured over time both the development of and the recovery from ILR, in the same manner as in experiment 1.

A. Method

1. Listeners

Twelve listeners, eight females, participated. Four had participated in previous experiments on ILR, including, for two of them, experiment 1. They met the same auditory criteria as for experiment 1. Their ages ranged from 21 to 32 years. Listeners were members of the laboratory, including the first author (L3), or undergraduates, who were paid for their participation.

2. Stimuli

All stimuli were 200 ms tone bursts, including raised-cosine rise/fall times of 6.67 ms. Six signal frequencies were tested: 0.5, 1, 2.5, 4, 6, and 8 kHz. As in experiment 1, the stronger inducer tone was presented at 80 dB SPL and the test tone at 70 dB SPL. The inducers and test tones were always at the same frequency.

3. Apparatus

The apparatus was the same as in experiment 1.

4. Procedure

The method of magnitude estimation was applied in the same manner as in experiment 1, i.e., listeners estimated loudness by assigning numbers to represent loudness magnitude. Judgments were made separately at each of the six frequencies, which were divided into two sets, one comprising 0.5, 2.5, and 6 kHz and the other 1, 4, and 8 kHz. Each set of frequencies was tested on a different day. This division ensured that within one session the frequencies tested were far enough apart so that tones at one frequency would not reduce the loudness of tones at other frequencies (cf. Marks and Warner, 1991). A session began with measurements of classical loudness functions at one set of frequencies (either 0.5, 2.5, and 6 kHz or 1, 4, and 8 kHz) and was followed by measurements of ILR at the other set of frequencies. For the classical measurements, tones at a given frequency were presented in random order at seven levels in 5 dB steps from 40 to 70 dB SPL. (Note that the level never exceeded that of the test tones presented subsequently in the same session.) The sequence of seven levels was repeated three times in a different order yielding three estimations at each level. The second and third estimations were averaged. These measurements serve to show how much loudness changes for a given change in level over a range of levels that exceeded the expected maximum amount of ILR.

The measurements of ILR, which came after the classical measurements, were made over more trials than in experiment 1, but were otherwise the same. In the first of two sessions, measurements were made at three frequencies from one of the two frequency sets described in the preceding paragraph. On each of the first two of a series of 83 trials, listeners heard a single 70 dB tone burst. They estimated the loudness of the burst on the second trial. On the third trial, an 80 dB inducer came on 1.8 s before a 70 dB test tone. Listeners estimated the loudness of the weaker test tone; they were to ignore the louder inducer. The ensuing sequence of inducers and test tones is described in the next section in conjunction with Table II. In the second session, the measurements were made at the three frequencies of the other frequency set not yet tested. The order of frequencies within a session was random, and the order of the two frequency sets over the two sessions was counterbalanced across listeners.

TABLE II. Geometric means of 12 estimations, one by each of the 12 listeners, at each frequency as a function of the time (and trial) at which the test tone was presented.

			Frequency (kHz)					
Trial	Time (s)	Inducer	0.5	1	2.5	4	6	8
2	**8**	**No**	**5.8**	**5.5**	**5.8**	**5.9**	**5.3**	**5.3**
3	12	Yes	5.1	5.1	5.3	5.6	4.9	5.0
8	32	Yes	4.5	4.6	4.5	4.8	4.0	4.7
13	52	Yes	3.9	4.5	4.4	5.0	4.0	4.5
18	72	Yes	4.2	4.4	4.0	4.8	4.2	4.3
23	92	Yes	4.2	4.1	4.1	4.5	3.9	4.2
28	112	Yes	3.9	4.7	3.9	4.5	3.9	4.2
33	132	Yes	3.7	4.3	3.8	4.7	3.9	4.0
38	152	Yes	3.9	4.2	3.6	4.7	3.6	4.1
43	172	Yes	3.8	4.3	3.9	4.6	3.7	3.8
48	192	Yes	3.5	4.2	3.5	4.4	3.6	4.1
53	212	Yes	3.7	4.2	3.6	4.2	3.4	4.0
58	232	Yes	3.4	4.0	3.8	4.6	3.5	3.7
63	252	Yes	3.5	3.6	3.9	4.2	3.6	3.6
68	**272**	**No**	**3.7**	**4.2**	**4.1**	**4.4**	**3.5**	**4.1**
73	**292**	**No**	**3.5**	**4.5**	**4.2**	**4.7**	**3.7**	**4.5**
78	**312**	**No**	**3.8**	**4.3**	**4.4**	**4.6**	**3.8**	**4.6**
83	**332**	**No**	**3.8**	**4.5**	**4.5**	**5.1**	**4.1**	**4.6**
Ratio of trial 2 to trial 63			1.65	1.51	1.50	1.40	1.48	1.47

B. Results and discussion

Table II gives the geometric means at the six frequencies as a function of the time (and trial) at which the test tone was presented. Each geometric mean is based on 12 magnitude estimations, one estimation by each of the 12 listeners. On trials 1 and 2, the 70 dB test tone was presented with no preceding inducer. On trials 3 through 63, the test tone was preceded on every trial by the 80 dB inducer. On trials 64 through 83, the test tone was again presented alone. Estimations were called for on trials 2 and 3 and every fifth trial thereafter. The last line of the table gives the ratio of the estimation on trial 2 (with no inducer) to that on trial 63 (the last trial with an inducer). This direct indication of the amount of loudness reduction varied from 28% (ratio of 1.40) to 40% (ratio of 1.65) across the six frequencies with a grand mean of 33% (ratio of 1.50). To examine the statistical

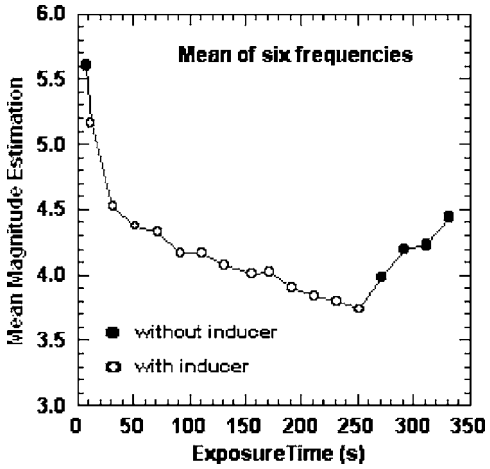


FIG. 2. Means of the original geometric means at all six frequencies as a function of exposure time, without (closed symbols) and with (open symbol) a preceding inducer.

significance of the effect of frequency ($n=6$) and listeners ($n=12$) on ILR, an ANOVA for repeated measures was performed. The ILR ratio of individual listeners was the dependent variable. The ANOVA showed no statistically significant effects for frequency (df=5, $F=0.43$, $p=0.8$). Differences among listeners were highly significant (df=11, $F=13.6$, $p<0.0001$).

As can be seen in Table II, neither the initial decrease in loudness over time nor its subsequent recovery depends in any consistent way on signal frequency. Consequently, it is reasonable to represent the change in loudness under ILR over exposure time by a single curve. The geometric means for all six frequencies were combined to yield the arithmetic means plotted in Fig. 2 as a function of exposure time. The first filled symbol and the last four are from trials with no inducer presented. The overall evolution of ILR over time is similar to that seen in Fig. 1, which shows the results from experiment 1 at a single frequency. Figure 2 shows clearly that loudness declines rapidly upon the introduction of the inducer at 12 s and continues to decline over at least the next 4 min; the rate of loudness decline tapers off after half a minute or so. However, at the time of the last inducer at 252 s on trial 63, the function appears not yet to have reached asymptote. On trial 68, 20 s after the removal of the inducer, the loudness of the test tone shows a small increase, indicating that recovery from ILR has begun. Loudness continues to increase gradually over the next 80 s, at which point the test ended, with recovery still not complete. After 80 s with no inducer, the estimated loudness is 4.4 as compared to an initial value of 5.6 prior to the introduction of the inducer.

Taken altogether, these results reveal no dependence of ILR on the signal frequency, at least not between 0.5 and 8 kHz. The decline in loudness averages 33% (a ratio of 1.50) after 4 min. On the assumption that the standard loudness function with an exponent of 0.6 applies (ISO R532, 1975), this reduction is equivalent to a decrease of 5.9 dB. An ILR of 5.9 dB is notably less than the 10 or 11 dB of ILR often measured. Another way to calculate the equivalent ILR in dB is to use, instead of the exponent 0.6 of the standard loudness function, the exponents measured in the "classical magnitude estimations" at the beginning of each of the two sessions of experiment 2. Based on the geometric means of the individual magnitude estimations as a function of the sound pressure level, those exponents vary from 0.23 to 0.31 across the six frequencies for a mean slope of 0.28, a low value but well within the range of published values (for a review, see Hellman, 1990). (Our low value probably reflects the listeners' restricting themselves to a small range of numbers, in many cases apparently ranking loudness.) The 33% loudness decline is then equivalent to a reduction of 12.6 dB. This overestimation of ILR is the corollary of the large *under*estimation of the slope of the loudness function. Such a flat slope means that a small change in the loudness of a tone corresponds to a large change in the level. This underestimation of the exponent would apply also to the judgments under ILR. However, the expected change in loudness under ILR was equivalent to no more than 10 or 11 dB. The underestimation over a range of 10 dB would be correspondingly less than that over the range of 30 dB to which these listeners were exposed in the classical magnitude estimations. Given that the use of the loudness exponents measured for these listeners leads to an estimation of ILR in equivalent dB that is too large (12.6 dB) and that the use of the standard exponent leads to an estimation that is too small (5.9 dB), it is not unreasonable to assume that the "true" amount of ILR at the six frequencies for this group of listeners is indeed reasonably close to the 10 dB usually measured by matching.

While the mean amount of ILR, taken as the ratio of the loudness before the introduction of an inducer tone and after 60 trials with an inducer tone, does not vary among frequencies (Table II), it does vary a great deal among listeners. Nonetheless, listeners generally are consistent in the amount of ILR they show, from frequency to frequency. The correlations (Pearson product moment) between the individual ILR ratios for pairs of frequencies were all statistically significant, ranging from 0.49 (0.5 kHz vs 6 kHz) to 0.99 (4 kHz vs 6 kHz) for a grand mean correlation of 0.76.

Finding no effect of frequency on ILR appears to be at variance with the one previous careful comparison of ILR on the same listeners at two frequencies, by Mapes-Riordan and Yost (1999), who used a 2AFC adaptive matching procedure. With the inducer at 80 dB SPL and the test at 70 dB SPL, they measured over 11 dB of ILR at 0.5 kHz and only 3 dB at 2.5 kHz. The 11 dB value is close to that reported by Nieder *et al.* (2003) who made measurements only at 0.5 kHz. However, the small value measured by Mapes-Riordan and Yost (1999) at 2.5 kHz is suspect. On the basis of similar loudness matches by eight listeners, Arieh and Marks (2003a) reported 11 dB of ILR at 2.5 kHz, a value close to that generally reported at 0.5 kHz. We also note that in another study, Mapes-Riordan and Yost (1998) reported under similar stimulus conditions 6 dB of ILR at 0.5 kHz and 5.5 dB at 2.5 kHz.

To shed some light on the discrepancy in the literature, loudness matches at 0.5 and 2.5 kHz and also at 8 kHz were performed—on separate days—by seven listeners who had served in experiment 2. The 2AFC adaptive matching procedure was identical to that used by Nieder *et al.* (2003). The

TABLE III. Amount of ILR in dB for seven listeners at three frequencies. (Arithmetic means and standard errors are given at the bottom. Measurements were made by a 2AFC matching procedure.)

	Frequency (kHz)		
	0.5	2.5	8
L 1	9.9	14.5	11.0
L 2	8.6	6.3	10.1
L 3	6.6	5.5	1.5
L 4	11.9	12.8	9.9
L 5	9.0	13.6	9.4
L 6	12.0	7.3	7.1
L 7	0.9	−1.0	6.4
Mean	8.4	8.4	7.9
SE	1.4	2.1	1.2

inducer tones and test tones always had the same frequency. The frequency of the comparison tone was 2.5 kHz to match a 0.5 kHz test tone, 0.5 kHz to match a 2.5 kHz test tone, and 4.5 kHz to match an 8 kHz test tone. As in the main part of experiment 2, the inducer tones were at 80 dB SPL, and the test tones were at 70 dB SPL. In agreement with the results from the magnitude estimations, the loudness matches reveal no dependence of ILR on signal frequency between 0.5 and 8 kHz. The mean amount of ILR is about the same at the three frequencies (Table III), with the average amount of ILR across frequencies equal to 8.2 dB. This amount of ILR is somewhat smaller than that reported in earlier studies based on loudness matches. The small amount reflects the failure of listener L7 to show any ILR at all at 0.5 and 2.5 kHz; without L7, the mean amount of ILR at those frequencies is close to the expected value of 10 dB.

IV. GENERAL DISCUSSION

The present results show clearly that the amount of ILR and its dependence on exposure time are about the same at frequencies from 0.5 to 8 kHz. Although limited to inducers at 80 dB SPL and test tones at 70 dB, we have little reason to believe that frequency would matter with other combinations of inducer and test levels. This conclusion agrees with most such data in the literature.

The results with respect to exposure time show that ILR begins already after a single exposure to an inducer. A similarly rapid onset is also suggested in the one other set of relevant published data, in Fig. 3 of Mapes-Riordan and Yost (1999). As to recovery from ILR after termination of the inducer, it begins quickly (Fig. 2) but then proceeds slowly. One minute after the last presentation of the inducer, the loudness of the test tone is still significantly below that of its initial value ($p<0.0001$) prior to the onset of the inducer. Although additional measurements on 4 of the 12 original listeners suggest that full recovery requires at least 10 min, and data reported by Mapes-Riordan and Yost (1998) suggest that full recovery would require as much as 20 min at 2.5 kHz, we note that Arieh *et al.* (2005) measured full recovery after little more than 2 min without an inducer. As suggested in the discussion of experiment 1, various differences in the stimulus conditions may account for the discrepancy. For example, recovery may be slower after longer-lasting exposures to inducers with longer durations as was the case in the present study and in that of Mapes-Riordan and Yost (1998).

The present findings lend further support to the notion posited in previous papers (e.g., Scharf, 2001; Nieder *et al.*, 2003) that ILR and ipsilaterally induced loudness adaptation (ILA) are based on the same underlying mechanism. Like ILR, ILA measures the effect of a stronger sound on the loudness of a weaker one. However, the duration and timing of the sounds are quite different. Under ILA, the weaker test tone is presented continuously and the stronger (inducer) tone is either an intermittent increment or an intermittent addition of a stronger tone at a different frequency (e.g., Canévet *et al.*, 1983; Charron and Botte, 1988). The stronger tone lasts at least 5 s and is usually repeated after every 10 to 60 s, depending on the study and condition. Despite these longer durations and the lack of a silent interval between inducer and test tones (but see Nieder *et al.*, 2003, p. 2852), the results for ILR and ILA are remarkably similar. Under ILA, the induced adaptation or loudness reduction has been measured mostly by magnitude estimation and is approximately 50%; under ILR, the reduction measured by loudness matching was also 50% (taken as the equivalent of the reported values of 10 dB). Moreover, the two kinds of loudness reduction depend in very similar ways on various stimulus parameters. The present results show that ILR, like ILA (Charron and Botte, 1988), is about the same over a wide range of stimulus frequencies. The present results also show that ILR, like ILA (Canévet *et al.*, 1985), develops quickly and then approaches asymptote after 2 or 3 min. (No data are available about the recovery from ipsilateral ILA.). Previous research showed that under both, the size of the loudness reduction decreases as the frequency separation between the inducer and test tones increases (Charron and Botte, 1988; Marks and Warner, 1991; Marks, 1994). Moreover, for both, the amount of loudness reduction is stable for individual listeners but varies considerably from one listener to another (Scharf *et al.*, 1983; Scharf *et al.*, 2003). All these similarities point to a common basic mechanism; this inference would be strengthened if it were shown that ILR and ILA are strongly correlated in individual listeners.

Decruitment may also be an example of how exposure to stronger stimulation results in a reduction of loudness. Canévet and Scharf (1990) found that when a sound decreases continuously in level, its loudness declines more rapidly than when it is presented discretely in quasirandom order over the same range of levels. Following Botte *et al.* (1986), they ascribed at least part of the rapid decline to self-induced adaptation, which is akin to ILA. However, unlike ILR, decruitment is strongest below about 40 dB SPL (Canévet and Scharf, 1990; Teghtsoonian *et al.*, 2000).

Still another paradigm for the measurement of the effect of a stronger sound on the loudness of a weaker one is related to the voluminous literature on the temporary threshold elevation or shift (TTS) induced by sound exposures that are often long and intense. The TTS is usually accompanied by a temporary decline in loudness often referred to as temporary loudness shift or TLS (e.g., Botte *et al.*, 1993). At moderate

exposure levels (Botte and Mönikheim, 1994), TLS resembles ILR. For example, 15 s after a 5 min exposure to a 65-dB-SPL tone at 1 kHz, the loudness of 1 kHz tone bursts at 50 and 60 dB SPL was down the equivalent of 10–11 dB; for bursts at 30 and 40 dB SPL, loudness was down approximately 8 dB. Like ILR, the amount of TLS was independent of signal frequency, but decreased when the test and exposure (inducer) frequencies were made different.

From all these considerations, it becomes clear that induced loudness reduction, in one guise or another, is a ubiquitous phenomenon. Generally, stronger tones reduce the loudness of weaker tones that follow closely in time (but not too closely). Does this generalization apply to sounds in general, to noises and multitone complexes, to time-varying sounds, to a succession of spectrally varying sounds, etc.? Must the stronger tone be presented many times to yield the persistent loudness reduction seen in the present studies? Does the reduction persist in the absence of continued auditory stimulation? These many unknowns make premature any attempt to uncover a specific mechanism for induced loudness reduction. They also leave uncertain the role of loudness reduction in everyday listening.

Nonetheless, given its persistence over time at all stimulus frequencies as the present experiments show, ILR must play a role in laboratory experiments, which so often use pure tones. Marks (1992) showed how ILR could alter loudness matches between distant frequencies as in the classical measurements by Fletcher and Munson (1933). Nieder *et al.* (2003) pointed out how ILR probably affected the outcome of measurements of the temporal integration of loudness. Unrecognized, ILR may well have played a role in many measurements of how loudness depends on signal level. Such measurements were usually based on magnitude estimation, with a series of tones presented over a range of levels in mixed order. Hence, stronger sounds would have at times reduced the loudness of succeeding weaker tones. The effect on the measured loudness function would depend on just how levels were ordered and how often and how long after a silent interval, tones were presented. Clearly, some of the variability among listeners and among studies would depend on these stimulus variables which were not thought to require careful monitoring. On the basis of experiments done in our laboratory showing how ILR depends on signal level, we can indicate how ILR is likely to affect the shape of the loudness function. Once we learn more about the kinds of sounds and stimulus conditions required for the manifestation of ILR, we shall be able to uncover its role in a multitude of psychophysical tasks and in listening, in general.

ACKNOWLEDGMENTS

Søren Buus was closely involved in the planning of these experiments. The authors thank Rhona Hellman, Michael Epstein, and Bob Teghtsoonian for many helpful comments about an earlier version of this paper. We also thank Alli Brower and Libby Gifford for making many of the measurements and Mary Florentine for her support. Yoav Arieh, an anonymous reviewer, and editor Andrew Oxenham made many very helpful suggestions. This research was supported by NIH/NIDCD Grant No. R01 DC 02241.

ANSI (1996). "Specification for audiometers," in *ANSI S3.6-1996* (American National Standards Institute, New York).

Arieh, Y., and Marks, L. E. (**2003a**). "Time course of loudness recalibration: implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Arieh, Y., and Marks, L. E. (**2003b**). "Recalibrating the auditory system: a speed- accuracy analysis of intensity perception," J. Exp. Psychol. Hum. Percept. Perform. **29**, 523–536.

Arieh, Y., Kelly, K., and Marks, L. E. (**2005**). "Tracking the time to recovery after induced loudness reduction," J. Acoust. Soc. Am. **117**, 3381–3384.

Botte, M. C., and Monikheim, S. (**1994**). "New data on the short-term effects of tone exposure," J. Acoust. Soc. Am. **95**, 2598–2605.

Botte, M. C., Baruch, C., and Scharf, B. (**1986**). "Loudness reduction and adaptation induced by a contralateral tone," J. Acoust. Soc. Am. **80**, 73–81.

Botte, M. C., Charron, S., and Bouayad, H. (**1993**). "Temporary threshold and loudness shifts: frequency patterns and correlations," J. Acoust. Soc. Am. **93**, 1524–1534.

Canévet, G., and Scharf, B. (**1990**). "The loudness of sounds that increase and decrease continuously in level," J. Acoust. Soc. Am. **88**, 2136–2142.

Canévet, G., Scharf, B., and Botte, M. C. (**1983**). "Loudness adaptation, when induced, is real," Br. J. Audiol. **17**, 49–57.

Canévet, G., Scharf, B., and Botte, M. C. (**1985**). "Simple and induced loudness adaptation," Z. Phys. D: At., Mol. Clusters **24**, 430–436.

Charron, S., and Botte, M. (**1988**). "Frequency selectivity in loudness adaption and auditory fatigue," J. Acoust. Soc. Am. **83**, 178–187.

Elmasian, R., and Galambos, R. (**1975**). "Loudness enhancement: monaural, binaural, and dichotic," J. Acoust. Soc. Am. **58**, 229–234.

Fletcher, H. F., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Hellman, R. P. (**1990**). "Loudness measurements by magnitude scaling: Implications for intensity coding," in *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S. S. Stevens*, edited by G. A. Gescheider and S. J. Bolanowski (Erlbaum, Hillsdale, NJ).

Hellman, R. P., Miskiewicz, A., and Scharf, B. (**1997**). "Loudness adaptation and excitation patterns: Effects of frequency and level," J. Acoust. Soc. Am. **101**, 2176–2185.

International Organization for Standardization. (**1975**). "Method for calculating loudness level," R532, New York.

Mapes-Riordan, D., and Yost, W. A. (**1998**). "Temporal properties of loudness recalibration," *Proc. 16th Int. Cong. Acoust.*, pp. 2361–2362.

Mapes-Riordan, D., and Yost, W. A. (**1999**). "Loudness recalibration as a function of level," J. Acoust. Soc. Am. **106**, 3506–3511.

Marks, L. E. (**1992**). "The contingency of perceptual processing: Context modifies equal-loudness relations," Psychol. Sci. **3**, 285–291.

Marks, L. E. (**1993**). "Contextual processing of multidimensional and unidimensional auditory stimuli," J. Exp. Psychol. Hum. Percept. Perform. **19**, 227–249.

Marks, L. E. (**1994**). ""Recalibrating" the auditory system: the perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.

Marks, L. E., and Warner, E. (**1991**). "Slippery context effect and critical bands," J. Exp. Psychol. Hum. Percept. Perform. **17**, 986–996.

Meiselman, H. L. (**1968**). "Magnitude estimations of the course of gustatory adaptation," Percept. Psychophys. **4**, 193–196.

Miskiewicz, A., Scharf, B., Hellman, R., and Meiselman, C. (**1993**). "Loudness adaptation at high frequencies," J. Acoust. Soc. Am. **94**, 1281–1286.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test- and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.

Parker, S., and Schneider, B. (**1994**). "The stimulus range effect: evidence for top-down control of sensory intensity in audition," Percept. Psychophys. **56**, 1–11.

Plack, C. J. (**1996**). "Loudness enhancement and intensity discrimination under forward and backward masking," J. Acoust. Soc. Am. **100**, 1024–1030.

Scharf, B., Botte, M. -C., and Canévet, G. (**1983**). "Loudness adaptation induced interaurally and monaurally," in *Hearing–Physiological Basis and Psychophysics*, edited by R. Klinke and R. Hartmann, (Springer-Verlag, New York), pp. 257–262.

Scharf, B. (**2001**). "Sequential effects in loudness," in *Fechner Day 2001: Proceedings of the 19th Annual Meeting of the International Society for*

Psychophysics, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Papst, Berlin), pp. 254–259.
Scharf, B., Buus, S., and Nieder, B. (**2002**). "Loudness enhancement: induced loudness reduction in disguise? (L)," J. Acoust. Soc. Am. **112**, 807–810.
Scharf, B., Nieder, B., Buus, S., and Cazals, Y. (**2003**). "Induced loudness reduction: Correlation between ears and procedures," in *Fechner Day 2003: Proceedings of the Nineteenth Annual Meeting of the International Society for Psychophysics*, edited by B. Berglund and E. Borg (ISP, Stockholm), pp. 281–286.
Scharf, B., Wagner, E., and Nieder, B. (**2005**). "Induced loudness reduction: A review," J. Acoust. Soc. Am. **117**, 2454(A).
Schlauch, R. S., Harvey, S., and Lanthier, N. (**1995**). "Intensity resolution and loudness in broadband noise," J. Acoust. Soc. Am. **98**, 1895–1902.
Teghtsoonian, R., Teghtsoonian, M., and Canévet, G. (**2000**). "The perception of waning signals: decruitment in loudness and perceived size," Percept. Psychophys. **62**, 637–646.
Zwicker, E. (**1958**). "Ueber psychologische und methodische Grundlagen der Lautheit" ("About psychological and methodical principles of loudness)", Acustica **8**, 237–258.

Comodulation detection differences for fixed-frequency and roved-frequency maskers

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

(Received 29 June 2005; revised 28 October 2005; accepted 18 November 2005)

This study investigated comodulation detection differences (CDD) for fixed- and roved-frequency maskers. The objective was to determine whether CDD could be accounted for better in terms of energetic masking or in terms of perceptual fusion/segregation related to comodulation. Roved-frequency maskers were used in order to minimize the role of energetic masking, allowing possible effects related to perceptual fusion/segregation to be revealed. The signals and maskers were composed of 30-Hz-wide noise bands. The signal was either comodulated with the masker (A/A condition) or had a temporal envelope that was independent (A/B condition). The masker was either gated synchronously with the signal or had a leading temporal fringe of 200 ms. In the fixed-frequency masker conditions, listeners with low A/A thresholds showed little masking release due to masker temporal fringe and had CDDs that could be accounted for by energetic masking. Listeners with higher A/A thresholds in the fixed-frequency masker conditions showed relatively large CDDs and large masking release due to a masker temporal fringe. The CDDs of these listeners may have arisen, at least in part, from processes related to perceptual segregation. Some listeners in the roved masker conditions also had large CDDs that appeared to be related to perceptual segregation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151788]

PACS number(s): 43.66.Dc, 43.66.Ba [GDK] Pages: 1021–1028

I. INTRODUCTION

There has long been interest in the possible role of comodulation among frequency-distributed spectral components in auditory perceptual organization (e.g., Broadbent and Ladefoged, 1957; Darwin, 1981; Bregman *et al.*, 1985). Such a role has been suggested for both comodulation masking release (CMR) (e.g., Hall *et al.*, 1984) and for comodulation detection differences (CDD) (Cohen and Schubert, 1987; McFadden, 1987; Wright, 1990). In the CDD paradigm, the detectability of a narrow band of noise depends upon the relation between its modulation pattern and the pattern(s) of modulation carried by spectrally distributed masking bands. When the signal and masking bands are comodulated (referred to here as the A/A condition), detection is relatively poor, but when the signal has a temporal envelope that is independent from the masker (referred to here as the A/B condition), detection improves. McFadden (1987) and Cohen and Schubert (1987) suggested that the good performance in the A/B condition arises because the across-frequency difference in modulation pattern allows the signal band to be segregated perceptually from the masker. By this interpretation, at least part of the masking in the A/A condition is due to a perceptual organization effect where detection is hampered by a difficulty in perceptually segregating comodulated noise stimuli. Recently, however, it has been suggested that CDD can be accounted for by peripheral auditory processes (Borrill and Moore, 2002; Moore and Borrill, 2002). Moore and Borrill (2002) pointed out that in the A/A condition, energetic masking is very efficient because the energy peaks of the masker coincide with the energy peaks of the signal; in the A/B condition, the energy peaks of the signal will sometimes coincide with the energy dips of the masker, making the masking less efficient. Moore and Borrill (2002) further pointed out that a contribution to masking based upon basilar-membrane suppression would have a similar effect; that is, in the A/A condition, the peaks of the masker would provide efficient suppression of the peaks of the signal, but in the A/B condition, some of the peaks of the signal would not coincide with masker peaks, and, thus, would not be suppressed as effectively. This argument assumes that the effects of suppression are virtually instantaneous (Arthur *et al.*, 1971).

Although the within-channel account of CDD given above is parsimonious and sound, it does not necessarily rule out the possibility that central processes related to perceptual fusion/segregation can contribute to CDD. The purpose of the present study was to explore this issue. The core of the approach was the use of informational masking involving masker frequency uncertainty (e.g., Neff and Callaghan, 1987; Neff and Green, 1987; e.g., Neff and Callaghan, 1988; Lutfi, 1990; Leek *et al.*, 1991; Kidd *et al.*, 1994; 2002; Richards *et al.*, 2002; Kidd *et al.*, 2003) within the CDD paradigm. A major benefit of this approach is that it limits the contribution of peripheral factors related to energetic masking/suppression, increasing the power to measure possible central contributions to observed differences between A/A and A/B masking conditions. In this approach, we varied the masker spectral composition on an interval-by-interval basis. In many listeners, this results in relatively high (poor) signal thresholds that are determined primarily by central factors (e.g., Neff and Callaghan, 1987; Kidd *et al.*,

[a]Electronic mail: jwh@med.unc.edu

1994). Kidd *et al.* (1994) showed that listeners can gain release from informational masking if cues are available that enable the signal to be segregated perceptually from the masker. Of particular interest here was whether across-frequency differences in temporal envelope would be associated with a release from informational masking.

One of the most important findings of Moore and Borrill (2002) pertained to conditions where the masker had a leading temporal fringe. They found that such a fringe had no effect on masked thresholds. This is highly significant from the standpoint of evaluating the hypothesis that one factor limiting the detection of the signal in the A/A condition is a difficulty in perceptually segregating the signal from its comodulated flanking bands. If such a perceptual segregation difficulty had been a material factor, then the segregation cue provided by a masker temporal fringe should have been associated with an improvement in the A/A threshold. The lack of such an effect was strong support for the hypothesis that CDD effects are driven by peripheral factors as opposed to factors associated with perceptual segregation. The current investigation employed a masker temporal fringe in some conditions in order to gain further insight into the possible role of cues related to perceptual segregation.

A general expectation in the present study was that individual differences might be prominent. McFadden and Wright (1990) found that individual differences were common in the CDD paradigm, with some listeners showing very low masked thresholds, and others showing considerably higher thresholds. Individual differences were also expected from the standpoint of informational masking, as it is often the case that some listeners show very little informational masking while others show informational masking effects of 40 dB or more (e.g., Kidd *et al.*, 1994; Neff and Dethlefs, 1995).

II. METHODS

A. Listeners

The listeners were two males and six females, ranging in age from 23 to 50 years. All listeners had previously participated in psychoacoustical tasks. Listeners had thresholds of 20 dB HL (ANSI, 2004) or better between octave frequencies of 0.25 and 8.0 kHz.

B. Stimuli

The masker/signal combinations were composed of sets of 30-Hz-wide bands of noise. The signal band was always centered on 2 kHz. The stimuli were played out of one channel of a real-time processor (RP2, TDT) at a rate of 24.4 kHz, routed to a headphone buffer (HB7, TDT) and presented over the left earphone of a pair of Sennheisser headphones (HD 265). There were two paradigms, one with fixed masker frequencies and the other in which the masker frequencies were roved. For each, the masker was composed of six noise bands. In the fixed-masker conditions, the center frequencies of the masking bands were 0.870, 1.169, 1.540, 2.570, 3.280, and 4.160 kHz. These frequencies were selected to leave approximately two nonoverlapping equivalent rectangular bandwidths for normal-hearing ears (ERB_N) between the edge frequencies of neighboring bands (Moore and Glasberg, 1983). In the roved-masker conditions, the masker frequencies on a given trial were chosen at random from a uniform distribution of log frequency, with the following restrictions: (1) the lower frequency limit was 0.25 kHz and the upper frequency limit was 8 kHz; (2) the edges of neighboring bands (including the signal band) were separated by at least two nonoverlapping ERB_N's. This criterion insured that the bands were no more closely spaced in the roved-masker conditions than in the fixed-masker conditions. In each paradigm, thresholds were obtained for both A/A (masking and signal bands all comodulated) and A/B (masking bands comodulated but the signal band random) stimuli. Each masker band was presented at a level of 48 dB SPL (a pressure spectrum level of approximately 33 dB/Hz).

In all conditions, stimuli were generated digitally based on arrays that were comprised of 2^{15} points, resulting in a total stimulus duration of 1.3 s prior to gating. Both the signal and masking bands were generated in the frequency domain based on Gaussian-distributed real and imaginary components. The same amplitude and phase draws were used to generate all bands in the A/A conditions. In the A/B conditions, the same draws were used to generate all masker bands, and independent draws were used to generate the signal band. The signal and the masker arrays were converted to the time domain via an inverse FFT.

The effect of signal/masker synchrony was investigated in both the fixed-masker and roved-masker conditions. The signal and masking bands were either gated synchronously, or in such a way that the masker bands had a 200-ms leading temporal fringe with respect to the signal band (all bands were coterminous). All gating was shaped via multiplication with a 50-ms raised cosine, and durations were defined with reference to the 50% rise points. In the temporal fringe conditions, both the signal and masking bands were gated on synchronously and had a total duration of 500 ms. In the asynchronous gating conditions, the maskers were gated on, the signal was gated on 200 ms later, and both were gated off 500 ms later, resulting in a total duration of 700 ms. The interstimulus interval was always 400 ms.

The use of the leading temporal fringe introduces an issue with respect to informational masking in the roved-masker condition. We have previously found order effects related to a masker temporal fringe in conditions involving masker frequency uncertainty (Hall *et al.*, 2005). Not only does such a fringe usually result in a large release from informational masking, but it can sometimes also be associated with a reduction in informational masking in a subsequent condition where the fringe is absent. Because this creates a potentially important order effect, the presentation order in the roved-masker conditions is a relevant consideration. The roved-masker conditions were therefore performed twice. For the first round, threshold estimates for a fringe-absent condition were run and then threshold estimates were obtained for the same condition but with the fringe present. This is referred to as the fringe-absent/fringe-present presentation order. After all roved-masker conditions were completed in this order, a second set of conditions was completed where threshold estimates for a fringe-present condition were

TABLE I. A/A and A/B thresholds (dB SPL) and derived measures of CDD and masking release (dB) due to masker fringe. Data are for the fixed-masker conditions. Standard deviations for the mean data are shown in parentheses.

	Fringe absent			Fringe present			Masking release due to fringe	
	A/A	A/B	CDD	A/A	A/B	CDD	A/A	A/B
L1	17.4	17.1	0.3	18.6	14.5	4.1	−1.1	2.6
L2	20.8	16.3	4.5	21.8	15.2	6.6	−1.0	1.1
L3	21.0	17.1	3.9	19.6	15.3	4.3	1.4	1.8
L4	27.2	22.4	4.8	27.0	21.1	5.9	0.2	1.3
L5	31.0	22.2	8.8	29.3	22.1	7.2	1.7	0.1
L6	34.9	15.7	19.2	23.4	14.1	9.3	11.5	1.6
L7	46.3	36.0	10.3	35.7	31.8	3.9	10.6	4.2
L8	48.2	31.6	16.6	35.9	31.3	4.6	12.3	0.2
Mean	30.8	22.3	8.5	26.4	20.7	5.7	4.4	1.6
	(11.6)	(7.6)	(6.6)	(6.8)	(7.4)	(1.9)	(5.9)	(1.3)

run, and then threshold estimates were obtained for the same condition but with the fringe absent. This is referred to as the fringe-present/fringe-absent order. One listener (L5) completed only the fringe-absent/fringe-present order. Extensive practice was provided to all listeners. The practice included completion of all of the conditions of the experiment and comprised approximately 50 threshold runs.

C. Procedure

Stimuli were presented in blocks of three-alternative, forced-choice trials, using a 2-down, 1-up adaptive tracking procedure estimating the 70.7%-correct point on the psychometric function (Levitt, 1971). The starting signal level was chosen such that the signal was clearly detectable. An initial step size of 4 dB was reduced to 2 dB after the second reversal. The run included eight reversals, with the signal threshold estimated as the average signal level at the final six reversals. A run typically consisted of 25–30 trials. Three to five such threshold estimates were obtained for each condition, with the final threshold taken as the mean of the estimates obtained.

III. RESULTS AND DISCUSSION

The average masked thresholds of the individual listeners are summarized in Table I and Fig. 1 (fixed-masker conditions), and in Tables II and III, and Fig. 2 (roved-masker conditions). Listeners are ordered (L1–L8) in terms of masked thresholds in the fixed-masker, fringe-absent, A/A condition (which ranged from 17.4–48.2 dB SPL). Before examining CDD effects in detail, it should be noted that there were large individual differences in the masked threshold data in terms of both the amount of masking in the fixed-masker conditions and the amount of informational masking in the roved-masker conditions. This is consistent with previous reports of substantial individual differences in susceptibility to masking in the CDD paradigm (e.g., McFadden and Wright, 1990) and in informational masking paradigms (e.g., Kidd *et al.*, 1994; Neff and Dethlefs, 1995). Although Fig. 2 shows that most listeners had high A/A thresholds in the roved-masker, fringe-absent conditions (indicating substantial informational masking), for two listeners (L2 and L4) thresholds in the roved-masker conditions were relatively low, indicating little informational masking. Listener L2 had demonstrated substantially higher roved-masker thresholds in the initial practice runs, but had apparently learned a strategy to overcome such masking by the time the final data were taken. Over all listeners, the A/A thresholds in the roved-masker conditions (temporal fringe absent) ranged from 18.8 to 74.3 dB SPL. The substantial variation in performance across listeners makes statistics based upon

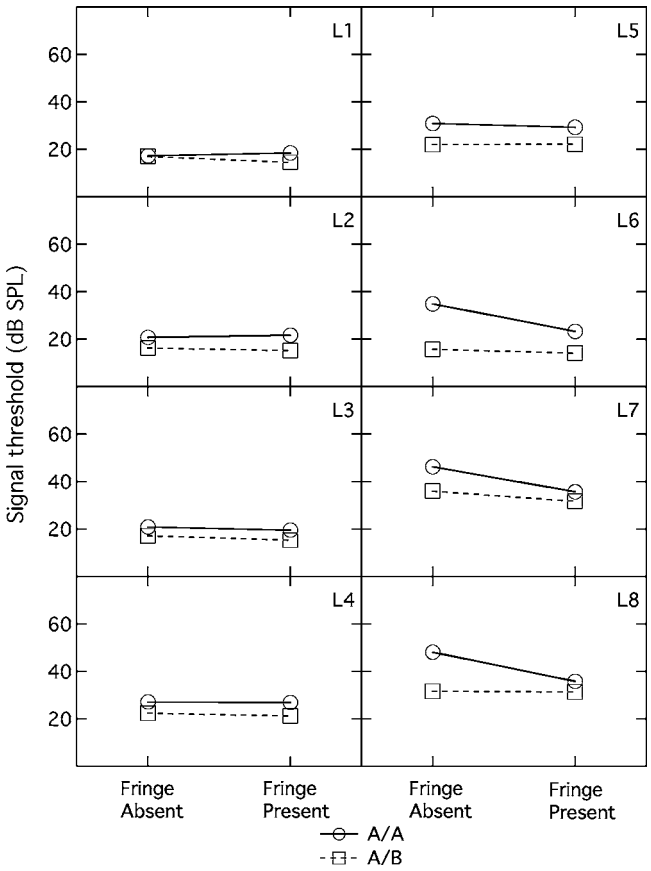


FIG. 1. A/A (circles) and A/B (squares) thresholds for individual listeners in the fixed-masker conditions.

TABLE II. A/A and A/B thresholds (dB SPL) and derived measures of CDD and masking release (dB) due to masker fringe. Data are for the roved-masker conditions for the presentation order fringe-absent/fringe-present. Standard deviations for the mean data are shown in parentheses.

	Fringe absent			Fringe present			Masking release due to fringe	
	A/A	A/B	CDD	A/A	A/B	CDD	A/A	A/B
L1	60.2	42.7	17.5	10.6	7.6	3.0	49.6	35.1
L2	22.9	15.5	7.4	13.6	9.7	3.9	9.3	5.8
L3	67.5	72.9	−5.4	12.1	8.3	3.8	55.4	64.6
L4	18.8	14.9	3.9	19.7	13.2	6.5	−0.8	1.7
L5	39.1	18.1	21.0	19.9	14.8	5.1	19.2	3.3
L6	65.2	39.1	26.1	14.9	12.2	2.7	50.3	26.9
L7	63.2	56.8	6.4	38.8	33.3	5.5	24.4	23.5
L8	74.3	71.0	3.3	35.8	25.8	10.0	38.5	45.2
Mean	51.4	41.4	10.0	20.7	15.6	5.0	30.7	25.8
	(21.4)	(24.0)	(10.5)	(10.8)	(9.2)	(2.4)	(20.8)	(22.3)

mean comparisons relatively uninformative. The analyses of data below therefore involve correlation and the examination of individual differences.

A. Fixed-masker conditions

Table I shows fixed-masker thresholds along with two kinds of derived measures, CDD and masking release resulting from temporal fringe. For fringe-absent conditions, CDD varied widely among listeners, ranging from 0.3 dB (L1) to 19.2 dB (L6). The average CDD across all listeners was 8.5 dB. Generally, the listeners with the higher A/A masked thresholds had larger CDDs ($r=0.79; p=0.02$). It was also the case that listeners with the higher A/A thresholds obtained relatively great masking release when a temporal fringe was available in this condition ($r=0.90; p=0.002$). This masking release due to temporal fringe ranged from −1.1 dB (L1) to 12.3 dB (L8) and averaged 4.4 dB. Some listeners also showed a masking release in the A/B condition when a temporal fringe was available. However, this masking release was small (1.6 dB on average). A CDD can also be calculated for the fringe-present A/A and A/B conditions. This CDD ranged from 4.1 dB (L1) to 9.3 dB (L6) and averaged 5.7 dB.

B. Roved-masker conditions

We will first consider conditions where the order of presentation was fringe-absent/fringe-present. Table II shows roved-masker thresholds along with the CDD and the temporal-fringe derived measures. As with the fixed-masker conditions, CDDs in the roved-masker condition with the temporal fringe absent varied widely across listeners, ranging from −5.4 dB (L3) to 26.1 dB (L6). The average CDD across all listeners was 10.0 dB. The masking release due to temporal fringe in the A/A condition ranged from −0.8 dB (L4) to 55.4 dB (L3) and averaged 30.7 dB. The correlation between the A/A masked threshold and CDD was not significant ($r=-0.12; p=0.98$), but the correlation between the A/A threshold and the masking release obtained when a temporal fringe was present was significant ($r=0.87; p=0.005$). The masking release due to temporal fringe in the A/B condition ranged from 1.7 dB (L4) to 64.6 dB (L3) and averaged 25.8 dB. Note that whereas the effect of temporal fringe in the A/B condition was relatively large, it was minimal in the previously described fixed-masker conditions. One reason why the A/B temporal fringe effect was relatively large in the roved-masker conditions is that, for some listeners, the across-frequency envelope cue available in the synchronous

TABLE III. A/A and A/B thresholds (dB SPL) and derived measures of CDD and masking release (dB) due to masker fringe. Data are for the roved-masker conditions for the presentation order fringe-present/fringe-absent. Standard deviations for the mean data are shown in parentheses.

	Fringe absent			Fringe present			Masking release due to fringe	
	A/A	A/B	CDD	A/A	A/B	CDD	A/A	A/B
L1	56.0	55.7	0.3	9.2	8.7	0.5	46.8	47.0
L2	28.2	8.5	19.7	11.4	6.7	4.7	16.8	1.8
L3	68.0	19.1	48.9	13.8	9.1	4.7	54.2	10.0
L4	20.2	19.3	0.9	20.8	15.7	5.1	−0.6	5.1
L6	45.5	16.7	28.8	20.2	17.7	2.5	25.3	−1.0
L7	59.6	51.5	8.1	37.9	37.7	0.2	21.7	13.8
L8	63.1	28.7	34.4	33.8	25.6	8.2	29.3	3.1
Mean	48.6	28.5	20.2	21.0	17.3	3.7	27.6	11.2
	(18.2)	(18.2)	(18.3)	(11.0)	(11.1)	(2.8)	(18.4)	(16.6)

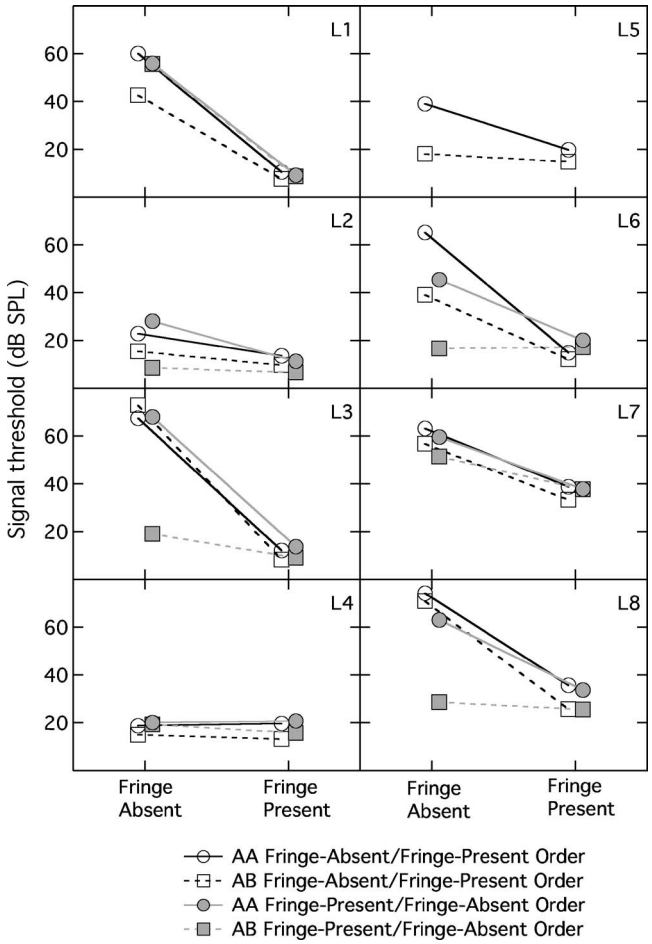


FIG. 2. A/A (circles) and A/B (squares) thresholds for individual listeners in the roved-masker conditions. The open symbols show data for the fringe-absent/fringe-present order of presentation and the gray symbols show data for the fringe-present/fringe-absent order of presentation.

A/B condition was not potent in achieving release from informational masking. This left a large potential range for masking release to occur when the temporal fringe cue was available. In line with this interpretation, the masking release obtained with temporal fringe in the A/B condition was highly correlated with the A/B masked threshold with fringe absent ($r=0.92$; $p=0.001$).

Table III shows roved-masker thresholds and derived measures for the fringe-present/fringe-absent presentation order. Again, CDD with the masker fringe absent varied widely across listeners, ranging from 0.3 dB (L1) to 48.9 dB (L3). The average CDD across all listeners was 20.2 dB. The masking release due to temporal fringe in the A/A condition ranged from −0.6 dB (L4) to 54.2 dB (L3) with an average masking release of 27.6 dB. The masking release due to temporal fringe in the A/B condition ranged from −1.0 dB (L6) to 47.0 dB (L1) and averaged 11.2 dB. The correlation between the A/A masked threshold and CDD was not significant ($r=0.51$; $p=0.25$). However, the correlation between the A/A threshold and the masking release obtained when a temporal fringe was present was significant ($r=0.82$; $p=0.02$).

Although there were similarities in the data patterns between the two orders of presentation (particularly for the fringe-present conditions), a comparison between Tables II and III reveals that there were individual cases where the A/B thresholds differed substantially between the two orders of presentation. Individual differences in the roved-masker conditions will now be examined in some detail.

C. Individual differences in roved-masker conditions

Two examples of individual variability within the roved-masker data set should be highlighted. The first involves listeners L3 and L8. Both of these listeners showed very large CDDs in the fringe-absent case, but only in the fringe-present/fringe-absent presentation order (cf. Tables II and III). For these listeners, the A/A thresholds with fringe absent were quite high for both orders of presentation (near 68 dB SPL for L3 and 63–74 dB SPL for L8), but the A/B threshold was much lower for the fringe-present/fringe-absent order (19.1 dB SPL for L3 and 28.7 dB SPL for L8) than for the fringe-absent/fringe-present order (over 70 dB SPL for both of these listeners). Because this effect of presentation order was remarkable, listener L8 was asked to repeat some of the roved-masker conditions approximately 1 month after completing the experiment. The listener first completed three A/B conditions in the order: fringe-absent, fringe-present, fringe-absent. This resulted in thresholds of 62.0, 28.4, and 31.6 dB SPL for these three conditions, respectively. This confirmed the previous finding that the listener performs poorly on the A/B fringe-absent condition, but improves on this condition immediately after listening to the A/B fringe-present condition. The A/A condition was then completed in the same order (fringe-absent, fringe-present, fringe-absent), and thresholds of 67.0, 34.3, and 68.6 dB SPL were obtained. This confirmed the previous finding that the listener performs poorly on the A/A fringe-absent condition, and does not improve on this condition after listening to the A/A fringe-present condition. The finding that a large CDD occurred for two listeners only when the A/B fringe-absent condition was preceded by the fringe-present condition is consistent with an interpretation that the across-frequency temporal envelope cue is useful to these listeners only after the fringe-present condition has primed the listeners to "hear out" the signal. For these listeners, the across-frequency envelope cue in the roved-masker condition is not particularly robust in isolation.

The other individual difference of note pertains to the roved-masker, fringe-absent data of listener L1. This listener showed a CDD of 17.5 dB in the fringe-absent/fringe-present order, but showed essentially no CDD in the fringe-present/fringe-absent order. As with L3 and L8, this effect was driven primarily by variability in the A/B threshold. Inspection of L1's practice data revealed that, whereas the A/A threshold was consistently high when fringe was absent (ranging from 60 to 73 dB SPL), the A/B threshold with fringe absent was highly inconsistent both within and between testing days (ranging from 17 to 68 dB SPL). This suggests that the across-frequency envelope cue was not sufficiently strong to make it reliably useful for this listener.

Thus, the results from L1 are also consistent with an interpretation that the across-frequency envelope cue in the roved-masker condition is not particularly robust. This interpretation is considered more generally in the following section.

D. Relative effectiveness of across-frequency envelope cues for release from informational masking

Several studies have shown that informational masking for roved-frequency maskers can be reduced substantially if cues are present that help the auditory system segregate the signal from the masker. For example, informational masking can be reduced by presenting the signal to one ear and the masker diotically (Kidd *et al.*, 1994). Here, informational masking is reduced by providing a spatial-hearing segregation cue. Informational masking can also be reduced by giving the masker a temporal fringe or a temporal pattern that is different from that of the signal (present study; Kidd *et al.*, 1994). Durlach *et al.* (2003) suggested that informational masking can be reduced by providing cues that make the signal sound different from the masker. The present A/B roved-masker condition can be considered a test of whether informational masking with roved-frequency maskers can be reduced by providing across-frequency differences in temporal envelope. The results suggest that this type of cue can provide release from informational masking. However, the cue does not appear to be as robust as the cue of temporal fringe. This question can be examined by considering the listeners who clearly showed informational masking in the roved-masker conditions (L1, L3, L5, L6, L7, and L8) and contrasting the informational masking release due to temporal envelope cues (the CDD for fringe-absent) and the masking release due to temporal fringe for the A/A stimulus. For the fringe-absent/fringe-present order, this contrast reveals that all six of the listeners with informational masking showed a masking release of more than 15 dB for the temporal fringe cue, but only three of six (L1, L5, and L6) showed masking release of more than 15 dB for the temporal envelope cue. For the fringe-present/fringe-absent order, this contrast reveals that all five of the listeners with high informational masking (L5 did not participate) showed a masking release of more than 15 dB for the temporal fringe cue, but only three of five (L3, L6, and L8) showed masking release of more than 15 dB for the temporal envelope cue. Thus, for the present listeners, across-frequency differences in temporal envelope were less reliable than temporal fringe in yielding substantial release (more than 15 dB) from informational masking.

IV. GENERAL DISCUSSION

A. Fixed-masker conditions

Borrill and Moore (2002) and Moore and Borrill (2002) suggested that CDD could be accounted for by peripheral factors (masking/suppression). The results from our listeners having the lowest masked thresholds are in good agreement with this idea. For example, whereas the five listeners having the lowest A/A thresholds (L1–L5) had an average CDD of approximately 5 dB, their average masking release due to temporal fringe in the A/A condition was only 0.2 dB. If their CDDs had been associated with a central mechanism related to perceptual segregation, a masking release due to temporal fringe would have been expected. The fixed-masker CDDs of these listeners are therefore consistent with a peripheral masking/suppression account.

It is likely that more central factors contributed to the CDDs of the present listeners who had higher masked thresholds in the fixed-masker A/A condition. The three listeners with the highest A/A thresholds (L6, L7, and L8) had CDDs of over 10 dB and masking release due to temporal fringe of over 10 dB. The observation that these listeners obtained a masking release with a temporal fringe is consistent with an interpretation that their A/A thresholds may not have been limited solely by peripheral factors. One possible interpretation is that the A/A thresholds of these listeners were limited by a failure to segregate the comodulated signal and masking bands (McFadden, 1987). By such a perceptual segregation account, the masking release associated with a temporal fringe should be greater for the A/A condition than for the A/B condition. The masking release for temporal fringe in the A/B condition would be smaller because the across-frequency temporal envelope cue would have already contributed to the perceptual segregation of the signal from the masker. For the three listeners with the highest A/A thresholds, the average masking release resulting from a temporal fringe was 11.5 dB for the A/A condition but only 2.0 dB for the A/B condition. This pattern of results is consistent with expectation based upon perceptual segregation.

Whereas three of our eight of listeners showed a substantial masking release due to temporal fringe in the fixed-masker, A/A condition, none of the three listeners tested in Moore and Borrill's (2002) conditions investigating temporal fringe showed such a masking release. Given the marked individual differences observed here and in previous data sets, it is possible that this discrepancy may be due entirely to subject sampling. Another possibility is that the difference in the pattern of results between the two studies is related to listener training. Although all of the listeners of the present study completed approximately 50 practice threshold runs on the conditions of this experiment before final data were collected, the listeners of Moore and Borrill (2002) may have been more highly practiced on CDD conditions. Moore and Borrill reported that their listeners had completed several previous CDD experiments in addition to receiving practice on the particular temporal fringe conditions completed. It is possible that perceptual fusion effects related to comodulation (as indicated by relatively poor A/A thresholds) can be reduced with extensive training.

B. Roved-masker conditions

The roved-masker paradigm was employed to increase the opportunity of observing central contributions to CDD. Some listeners showed very high thresholds in the A/A fringe-absent condition, presumably due to informational

rather than energetic masking. In some of the listeners, the A/A threshold for the signal band was higher in level than the levels of the individual masker bands (48 dB SPL). This was true for L1, L3, L6, L7, and L8 in the fringe-absent/fringe-present order and for L1, L3, L7, and L8 in the fringe-present/fringe-absent order. The CDDs of these listeners ranged from −5.4 to 51.5 dB, averaging 9.6 dB in the fringe-absent/fringe-present order and 32.3 dB in the fringe-present/fringe-absent order. It is very likely that the A/A thresholds of these listeners were determined primarily by central rather than peripheral factors. For example, because the signal band was well above the level of the masking bands at masked threshold, it is very unlikely that the flanking bands suppressed the signal band. In general, the roved-masker results are consistent with the idea that sensitivity to across-frequency differences in temporal envelope can result in masking release (CDD), although less robustly than for the segregation cue of masker temporal fringe.

As noted above, there were many individual differences in the roved-masker paradigm, including differences in susceptibility to informational masking, and in the ability to benefit from cues associated with across-frequency differences in temporal envelope. One possible source of variation in informational masking studies such as this may be an individual's ability to monitor a variety of detection cues throughout the course of an adaptive threshold run. For example, whereas all listeners may be able to use an overall energy cue to detect a signal in a roved masker when the signal-to-noise ratio is high, listeners may vary in their ability to switch to a different cue (e.g., across-frequency difference in temporal envelope) as the signal-to-noise ratio decreases. It is possible that the ability to switch among cues as the signal-to-noise ratio changes contributes to the individual differences found in complex listening tasks.

V. CONCLUSIONS

(1) In fixed-masker conditions, the CDDs of some listeners are compatible with an interpretation based upon peripheral masking/suppression. These listeners had relatively low masked thresholds in the A/A condition, and had little or no masking release related to a masker temporal fringe.
(2) Other listeners had relatively high A/A thresholds in the fixed-masker conditions and had relatively larger CDDs in the fringe-absent conditions. These listeners showed substantial masking release related to a temporal fringe in the A/A condition, but considerably less for the A/B condition. These results are consistent with an interpretation that at least part of the CDD in these listeners arises from across-frequency envelope cues that help the listener segregate the signal from the masker.
(3) In the roved-masker conditions, the results of several listeners were again compatible with an interpretation that CDD can result from across-frequency temporal envelope cues that help the listener segregate the signal from the masker. Here, substantial CDDs sometimes occurred in cases where peripheral masking and suppression were very unlikely to have contributed significantly to the A/A thresholds.
(4) Several aspects of the results indicate that release from informational masking related to across-frequency differences in temporal envelope is not as robust as informational masking release related to temporal onset asynchrony.

ACKNOWLEDGMENTS

We thank Gerald Kidd, Brian Moore, and an anonymous reviewer, for helpful comments on a previous version of this manuscript. This work was supported by NIH NIDCD Grant R01 DC00418. We thank Heidi Reklis and Madhu B. Dev for assistance in running subjects and for technical support.

ANSI (**2004**). ANSI S3.6-2004, "Specification for audiometers" (American National Standards Institute, New York).
Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). "Properties of 'two-tone inhibition' in primary auditory neurones," J. Physiol. (London) **212**, 593–609.
Borrill, S. J., and Moore, B. C. (**2002**). "Evidence that comodulation detection differences depend on within-channel mechanisms," J. Acoust. Soc. Am. **111**, 309–319.
Bregman, A. S., Abramson, J., Doehring, P., and Darwin, C. J. (**1985**). "Spectral integration based on common amplitude modulation," Percept. Psychophys. **37**, 483–493.
Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.
Cohen, M. F., and Schubert, E. D. (**1987**). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **81**, 721–723.
Darwin, C. J. (**1981**). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. A **33A**, 185–207.
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.
Hall, J. W., Buss, E., and Grose, J. H. (**2005**). "Informational masking release in children and adults," J. Acoust. Soc. Am. **118**, 1605–1613.
Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.
Kidd, G., Jr., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2002**). "Informational masking in listeners with sensorineural hearing loss," J. Assoc. Res. Otolaryngol. **3**, 107–119.
Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.
Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Leek, M. R., Brown, M. E., and Dorman, M. F. (**1991**). "Informational masking and auditory attention," Percept. Psychophys. **50**, 205–214.
Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lutfi, R. A. (**1990**). "How much masking is informational masking," J. Acoust. Soc. Am. **88**, 2607–2610.
McFadden, D. (**1987**). "Comodulation detection differences using noise-band signals," J. Acoust. Soc. Am. **81**, 1519–1527.
McFadden, D., and Wright, B. (**1990**). "Temporal decline of masking and comodulation detection differences," J. Acoust. Soc. Am. **88**, 711–724.
Moore, B. C., and Borrill, S. J. (**2002**). "Tests of a within-channel account of comodulation detection differences," J. Acoust. Soc. Am. **112**, 2099–2109.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Neff, D. L., and Callaghan, B. P. (**1987**). "Simultaneous masking by small

numbers of sinusoids under conditions of uncertainty," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Lawrence Erlbaum Associates, Hillsdale, N.J.).
Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.
Neff, D. L., and Dethlefs, T. M. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.
Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.
Richards, V. M., Tang, Z., and Kidd, G. D., Jr. (**2002**). "Informational masking with small set sizes," J. Acoust. Soc. Am. **111**, 1359–1366.
Wright, B. A. (**1990**). "Comodulation detection differences with multiple signal bands," J. Acoust. Soc. Am. **87**, 293–303.

Resolution in azimuth sound localization in the Mongolian gerbil (*Meriones unguiculatus*)

Julia K. Maier and Georg M. Klump[a)]
Zoophysiology and Behavior Group, Institute for Biology and Environmental Science, Carl von Ossietzky University, Postfach 2503, 26111 Oldenburg, Germany

(Received 9 May 2005; revised 16 October 2005; accepted 28 November 2005)

Minimum resolvable angles (MRAs) for sound localization in azimuth in the gerbil were determined in a behavioral study using tones, 300-Hz bands of noise centered at frequencies between 500 Hz and 8 kHz and broad-band noise of on average 60 dB SPL overall level. Using the method of constant stimuli, seven gerbils were trained in a two-alternative-forced-choice procedure to indicate if sounds were presented to them from the left or from the right by choosing the left or right arm of a Y-shaped cage. The MRA is the minimum angle between two loudspeaker locations that the gerbils discriminated. Animals were either stimulated from the front ($N=4$) or from the back ($N=3$). The MRA for broad-band noise randomly varying in level by ±6 dB was 23° and 45° for gerbils stimulated from the front or back, respectively. Generally a gerbil's MRA for tones declined up to 2 kHz reaching 20° and 31° for gerbils stimulated from the front or back, respectively, and the MRA was generally increased above this frequency. Results for narrow-band noise stimuli were similar. Results are discussed with respect to the available interaural cues and physiological mechanisms of sound localization in the gerbil. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159429]

PACS number(s): 43.66.Qp, 43.66.Pn, 43.66.Gf [GMK] Pages: 1029–1036

I. INTRODUCTION

According to the duplex theory of sound localization (Rayleigh, 1907), interaural differences in the time of arrival and in the level of the sound reaching the ear are used to compute the azimuth position of the sound source. For high-frequency sounds, interaural level differences (ILDs), and for low-frequency sounds, interaural time differences (ITDs), are thought to provide information about the azimuthal position of the sound source, although also envelope fluctuations can be exploited at high frequencies by the auditory system to estimate ITDs (see review by Yin, 2002). Jeffress (1948) proposed that ITD processing in the auditory system is based on an array of coincidence-detector neurons receiving excitatory input from both ears by means of delay lines—axons of different path length. Each neuron is excited maximally if the conduction delay resulting from the difference in path length compensates for the ITD of the sound stimulus, with the consequence that action potentials from each ear arrive simultaneously. Thus, the physiological range of ITDs that results from the physical dimensions of the head is represented by a neuronal place code, i.e., the range of azimuthal sound positions is mapped to an array of neurons responding maximally to different ITDs. Such a map of ITDs is found in *nucleus laminaris* of the barn owl (Carr and Konishi, 1988) and the chicken (Overholt *et al.*, 1992), providing support for the Jeffress model. However, recent investigations of ITD processing in the Mongolian gerbil medial superior olive (MSO, i.e., the mammalian equivalent to *nucleus laminaris*) suggest a pattern of ITD coding that deviates from the Jeffress model (Brand *et al.*, 2002). In the gerbil, a species with well-developed low-frequency hearing (thresholds are similar to those of humans in the range from 1 to 10 kHz, see Ryan, 1976), the ITDs to which neurons in the MSO respond maximally lie beyond the physiological range. Within the physiological range, the neurons' response functions show steep slopes. This is consistent with different ITDs being represented in the form of a rate code, rather than by different neurons responding maximally to a specific ITD. Since the gerbil has become a key species for understanding mechanisms of azimuth sound localization in mammals, it is desirable to evaluate its resolution in azimuth sound localization behaviorally. To date, Heffner and Heffner (1988a), have determined the Gerbil minimum azimuth angle necessary to discriminate between two sound-source locations using bursts of broadband noise. Such data, however, are lacking for other types of signals. Here we present additional results on the minimum angle for sound source separation [minimum resolvable angle (MRA)] for tones and bursts of narrow-band noise that may allow for a better evaluation of the performance of the gerbil auditory system at different sound frequencies. The MRA is the angular separation between the locations of two sound sources that an experimental subject has to discriminate. In this discrimination the experimental subject has to designate a single source broadcasting a signal. By analyzing the performance in this absolute localization task in relation to the separation of the loudspeakers, it is possible to assess the binaural cues available to the subjects (e.g., see Klump *et al.*, 1986).

[a)]Author to whom correspondence should be addressed. Electronic mail: georg.klump@uni-oldenburg.de

II. MATERIAL AND METHODS

A. Subjects

Seven adult Mongolian gerbils (*Meriones unguiculatus*), 3 females and 4 males, were used in this experiment. They originated from two strains bred at Regensburg, Germany (the laboratory of Otto Gleich) and Munich, Germany (the laboratory of Benedikt Grothe). Both were derived originally from the Tumblebrook strain. Their age at testing ranged between 6 and 9 months. The animals were housed in individual cages ($42\times26\times15$ cm) with a layer of litter (Raiffeisen) and slices of papers for burrowing. All cages were stored in a ventilated cage rack (Tecniplast Slim Line Sealsafe). All animals had unrestricted access to water. Body weight during experiments was measured daily to maintain the gerbils at an average 82.6% of their ad libitum weight of 71.4 ± 13.8 g (mean±SD). The experiments involving the deprivation took an average of four months including the training and the testing phase. Food rewards during experiments consisted of 20 mg pellets (Bioserve: Dustless Precision Pellets, Formula #FO163). Where necessary, additional rodent pellets (Altromin 1314) were fed to maintain the animals' weight roughly constant. The care and treatment of the animals was approved by the Bezirksregierung Weser-Ems (Permit No. 509.8-42502-35/7) and followed the NIH guide for the care and use of laboratory animals.

B. Apparatus

Experiments were performed in a sound-attenuating chamber (Industrial Acoustics type IAC 1203 A, with internal dimensions $224\times214\times198$ cm). The inside walls of the chamber were lined with a layer of sound-absorbing wedges (Illbruck Illsonic Pyramide 100/100, mounted on Illbruck Plano type 50/0 SF). The wedges had an absorption coefficient of more than 0.99 for frequencies above 500 Hz. A 10-cm layer of polyurethane foam (25 kg/m^3) lined the floor.

Experiments were performed in a Y-shaped cage constructed from wire mesh that was 10 cm high. The base of the Y consisted of an entrance area (10×10 cm) that led to a platform (8×6 cm) where the gerbil sat and listened to stimulus presentations. From the listening platform, the gerbil could run into the left and right arms of the Y (each 25×10 cm and separated by an angle of 90°) to indicate its response [Fig. 1]. The cage was fixed on top of a chipboard platform ($50\times40\times1.6$ cm) covered with a layer of sound absorbing wedges (Illbruck Illsonic Waffel 70/125, mounted on Illbruck Plano type 50/0 SF). A light-interrupting switch was mounted in each compartment (entrance area, platform, left and right arms) to register an animal's whereabouts. A feeding dish was mounted at the end of each of the two arms opposite to the listening platform. A feeder LED between the two arms served as an additional reinforcer. A platform LED was also placed between the two arms that was switched on during a trial. Two custom-built feeders mounted about 60 cm above the experimental cage were connected to a feeder dish via a flexible tube and dispensed the reward pellets. Three pairs of loudspeakers (Canton Plus XS) were mounted on a second chipboard ($115\times65\times1.6$ cm), which was also covered with a layer of sound absorbing wedges (Illbruck Illsonic Waffel 70/125, mounted on Illbruck Plano type 50/0 SF). The angle between the pairs of speakers on either side of the midline of the setup was fixed at 10°, 30°, or 60° azimuth as determined from the position of the gerbil on the platform at a distance of 85 cm. Sounds were presented from the front ($N=4$ animals facing the speaker) or from the back ($N=3$ animals turning their backs toward the speaker) of the gerbil sitting on the platform. Stimulating subjects from the back served to determine whether the ability to localize sound sources in the frontal hemi-field, which includes the field of best vision, is better then in the caudal hemi-field. Both chipboards were located on a pedestal constructed of metal bars 55 cm above the floor.

C. Stimulus generation

All stimuli were generated using a Linux workstation (AMD Processor, Sound Blaster PCI 512, 44.1 kHz sampling rate). Signals were attenuated using a programmable attenuator (PA4, Tucker Davis Technologies) before being amplified (Yamaha Stereo Amplifier A-520) and equalized (Technics Stereo Graphic Equalizer SH-8075). After equalization, any two speakers differed by less than 7.7 dB in the frequency range from 0.25 to 12 kHz, and the frequencies in the spectrum of a broad-band noise (0.2–12 kHz) varied by less than ±6.3 dB. A Power Multiplexer (PM1, Tucker Davis Technologies) enabled one of the six loudspeakers in the chamber to be activated upon signal presentation. By eliminating all dc offsets in the apparatus and switching when no signal was presented we made sure that no switching transients were available as unwarranted cues. The sound-pressure level (SPL) in the experimental setup was calibrated every morning with a sound level meter (Brüel & Kjær 2238 Mediator) with the microphone located at the position on the platform where the animal's head would normally be during the experiment just prior to stimulus presentation.

Three types of stimuli were employed, (1) sine waves with frequencies of 0.5, 1, 1.5, 2, 4, or 8 kHz, (2) bands of noise 300 Hz wide centered at 0.5, 1, 1.5, 2, 4, or 8 kHz, and (3) broad-band noise, limited by the transmission characteristic of the speakers (130 Hz–21 kHz). All stimuli were 125 ms in duration including 25-ms Hanning ramps at stimulus onset and offset. Stimuli were presented with a mean overall level of 60 dB SPL that was randomly varied by ±6 dB. Only the broad-band noise was presented with both constant and varying level. 60 dB SPL corresponds to an average sensation level of 41, 54, 55, 57, 57, and 55 dB for frequencies of 0.5, 1, 1.5, 2, 4, and 8 kHz, respectively. These calculations are based on threshold data from Ryan (1976).

D. Procedure

Animals were trained by means of operant conditioning in a two-alternative-forced-choice (2-AFC) paradigm with food rewards. Gerbils initialized a variable waiting interval of between 1 and 7 s upon jumping onto the platform. At this point, the platform LED was illuminated to signal to the animal the beginning of a new trial. After the waiting interval

was completed, a single stimulus was presented from one of the loudspeakers. Gerbils were trained to leave the platform and indicate if the stimulus was presented from the left or the right by entering the appropriate response compartment on the left or right side. If the gerbil responded correctly, the feeder LED was illuminated and the feeder on the appropriate side delivered a reward with a probability of 80%. Following food delivery, the next trial was initiated. If the wrong response compartment was chosen, the platform LED was switched off and the next trial could not be initiated until a period of 3 s had elapsed. If a stimulus trial was missed, i.e., if the animal remained on the platform, the platform LED was switched off and a new trial initiated after 3 s. If the subject missed a stimulus by leaving the platform but not breaking the light-interrupting switch in one of the response compartments, the trial ended after 5 s and the platform LED was switched off. The experimental protocol was monitored by the workstation using a custom program.

Discrimination thresholds for deriving the MRA were determined using the method of constant stimuli. Sessions were comprised of 11 blocks of 6 trials. In each block each of the loudspeakers was randomly activated. The first 6 trials of each experimental session served as a "warm-up" period, during which only stimuli from the widest angle (60°) were presented. It was discarded from the analysis. Each session lasted between 20 and 30 min.

E. Training

Initially, gerbils were trained using speakers with an angular separation of 120°, as determined from the position of the gerbil waiting on the platform. The stimulus duration was fixed at 800 ms. For determination of thresholds, stimulus duration was reduced to 125 ms. A single-frame video analysis of gerbils waiting on the response platform revealed that the signal duration of 125 ms was sufficiently short to assure that the gerbils were forced to perform under open-loop conditions. In the open-loop condition the animal is not able to turn its head in response to the activated speaker and thus adjust its head position in an orientation response to change the binaural cues available for localization. The average reaction times, i.e., the time between stimulus onset and arrival in the light interrupting switch in one of the arms of the y-maze, was 1.3 and 1.8 s for gerbils facing the speakers and gerbils turning their back toward the speakers, respectively. It usually took more than 400 ms after signal onset before the start of an oriented movement off the platform could be observed.

F. Determining the minimum resolvable angle

Data from three sessions (180 trials) were combined for determination of the MRA and psychometric functions were constructed combining stimulus presentation from left and right by plotting percent correct responses versus the angle between speaker pairs. In these three sessions the animals had to respond to at least 20 of the 30 stimuli, presented via a certain speaker, otherwise a fourth session was conducted and the first session was discarded from the analysis. Experimental sessions were excluded from the analysis if the animal responded less than 75% correct to stimuli presented via speakers, which enclosed the widest angle (60°) and if the animal did respond to less than 75% of all presented stimuli. The MRA was obtained by means of linear interpolation of the data in the psychometric function as the angle resulting in 75% correct responses.

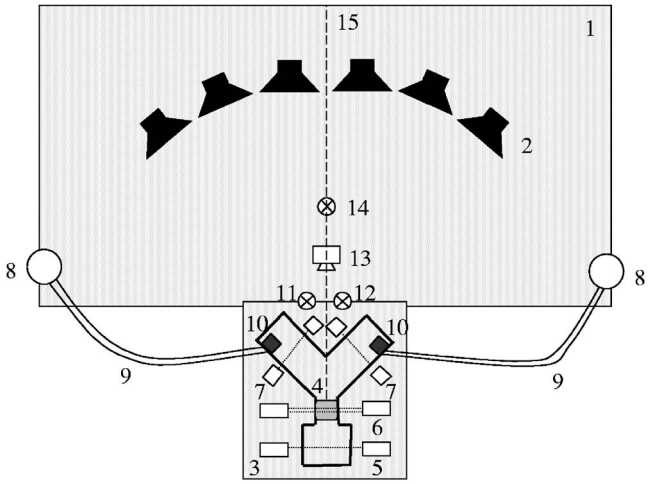


FIG. 1. Apparatus for the behavioral experiments. Large flake board (1) with loudspeakers (2) and small (3) flake board with cage (4). Light interrupting switch at the entrance area (5), double light interrupting switch at the platform (6), left and right light interrupting switches (7), feeders (8), flexible tubes (9), feeder dishes (10) feeder LED (11), platform LED (12), camera (13)*, roomlight (14)*, (15) symbolizes the 0°-line (* mounted 0.6 m above the flake board).

III. RESULTS

Initial training, performed using a pair of speakers separated by an angle of 120° and using a signal duration of 800 ms, took 18.6±5.8 (mean±SD) days. After this time, the minimum azimuthal angle each animal could resolve, the MRA was determined using stimuli of 125-ms duration. Not all experimental sessions could be used in determining the MRA: in 21.3% of sessions animals responded with fewer than 75% correct responses to stimuli presented via speakers that were separated by the widest angle (i.e., 60°), in 0.8% of sessions animals left the platform on fewer than 75% of the trials (irrespective of whether they chose the correct arm), in 3.3% of the sessions both criteria were violated. In three cases data from a fourth session had to be collected to obtain responses from at least 20 stimulus presentations from each speaker in three sessions; the first session of the four was discarded. The determination of a gerbil's MRA required, on average, 4.4±2.2 (mean±SD) sessions (Fig. 1).

A. Broad-band noise

Gerbils discriminated well between broad-band noise originating from speakers in the front that were separated by 60° [Fig. 2(a)]. On average, animals chose the correct side in 94% of all trials (range 90%–98%) when locating a broad-band noise with a level that varied randomly by ±6 dB. Maintaining the level at 60 dB SPL did not improve performance (average 94% correct, range 88%–100%). MRA performance (i.e., 75% correct) was reached at much smaller

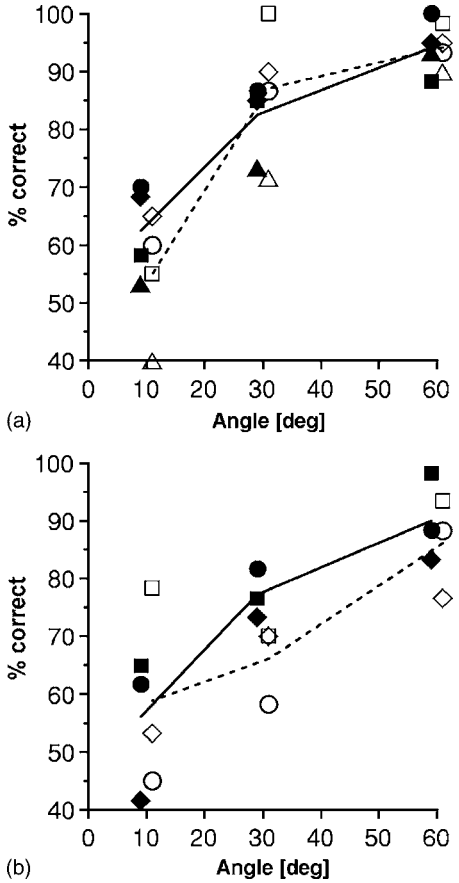


FIG. 2. Mean performance (% correct) for the localization of broad-band noise in relation to the angle between the speakers for stimulation from the front [(a), $N=4$] and from the back [(b), $N=3$]. Data from individual gerbils are identified by the differently shaped symbols. Closed symbols indicate data obtained with fixed-level noise bursts, open symbols indicate data obtained with noise bursts varying in level by ±6 dB. For lucidity closed symbols are plotted at an angles of 9°, 29°, and 59°, open symbols at angles of 11°, 31°, and 61°. Solid and dashed lines indicate mean performance for fixed and varying level, respectively.

angles. Average performance in response to broad-band noise presented from speakers separated by an angle of 30° was 87% and 83% for stimuli with varying and fixed level, respectively. Only at an angle of 10° did average performance fall below the MRA criterion (i.e., to 55% and 63% for stimuli with varying and fixed level, respectively). The average MRA across animals, when stimulated from the front, was 22.3° and 23.4° for fixed and randomized-level broad-band noise, respectively. The auditory resolution in azimuth was slightly reduced in subjects stimulated from the back [Fig. 2(b)] for localizing broad-band noise with fixed level (threshold 28.5°). For broad-band noise presented with randomized level, a significant increase in MRA was observed (45.2°, Tukey Test, $P<0.01$, following a two-way repeated-measures ANOVA with stimulus type and the direction from which the stimuli impinged on the animal as factors indicating significant effects of stimulus type, direction, and their interaction, all $p<0.05$) in animals stimulated from the back compared to those stimulated from the front. MRAs obtained in response to broad-band noise were among the lowest MRAs measured for the different stimuli.

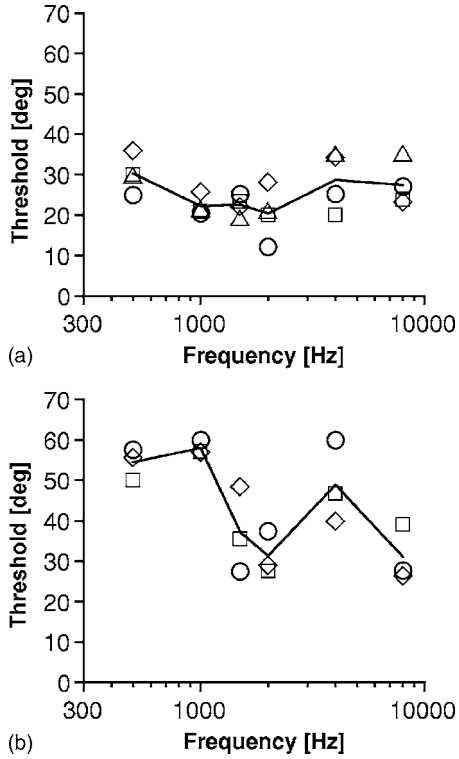


FIG. 3. Minimum resolvable angle (MRA) for the localization of tones in relation to the tone frequency: (a) data for stimulation from the front ($N=4$), (b) data for stimulation from the back ($N=3$). Data from individual gerbils are identified by the differently shaped symbols. The lines indicate the mean data.

B. Tones

Gerbils discriminated well between sounds originating from speakers in the front that were separated by the largest angle of 60°. The average performance levels for gerbils stimulated from the front at this angle were 85.4%, 94.2%, 93.8%, 92.5%, 90.4%, and 91.3% correct responses for frequencies of 0.5, 1, 1.5, 2, 4, and 8 kHz, respectively. With decreasing speaker separation, performance was significantly reduced. Gerbils exhibited the smallest MRA for tones between 1 and 2 kHz [mean MRAs ranged from 20.4° to 22.2°; Fig. 3(a)]. The MRAs at frequencies below and above this frequency range were slightly increased. The MRA was 30.3° for 500-Hz tones, 28.7° for 4-kHz tones, and 27.4° for 8-kHz tones. These differences were not significant (Tukey test, $P>0.05$, although a two-way repeated-measures ANOVA with stimulus type and the direction from which the stimuli impinged on the animal as factors in general indicated significant effects of stimulus type, direction, and their interaction, all $p\leqslant 0.001$). Gerbils stimulated from the back responded correctly to 80.0%, 76.1%, 87.2%, 89.4%, 80.6%, and 85.0% of the stimuli for frequencies of 0.5, 1, 1.5, 2, 4, and 8 kHz, respectively, at an angle separation of 60°. Animals stimulated with tones from the back showed significantly higher MRAs than animals stimulated from the front [Fig. 3(b); Tukey Test,$P<0.025$], with the exception of 8 kHz ($P=0.433$). In addition a much larger variation in MRA was observed (means ranging between 31.1° and 58.0°) for stimulation from the back as compared to the front. The lowest MRAs were observed for 2-kHz (31.4°)

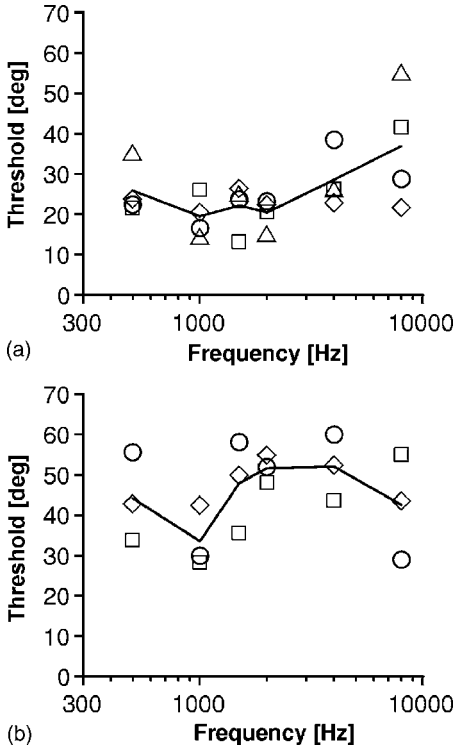


FIG. 4. MRA for the localization of narrow-band noise in relation to the center frequency of the noise band: (a) data for stimulation from the front (N=4), (b) data for stimulation from the back (N=3). Data from individual gerbils are identified by the differently shaped symbols. The lines indicate the mean data.

and 8-kHz tones (31.1°). Significant MRA differences were observed between some frequencies [0.5 vs 1.5 kHz, 0.5 vs 2 kHz, 0.5 vs 8 kHz, 1 vs 1.5 kHz, 1 vs 2 kHz, 1 vs 8 kHz; 2 vs 4 kHz, 4 vs 8 kHz, Tukey Test, all $P<0.025$, see Fig. 3(b)].

C. Narrow-band noise

Gerbils stimulated from the front with narrow-band noise performed at an angle separation of 60° correctly to 89.6%, 96.3%, 91.7%, 93.3%, 92.9%, and 88.8% of the presented stimuli for frequencies of 0.5, 1, 1.5, 2, 4, and 8 kHz, respectively. Similar to the localization of tones, gerbils showed the lowest MRA values for noise bands centered at frequencies between 1 and 2 kHz [variation in mean values from 19.5° to 20.4°; Fig. 4(a)]. MRAs at all other center frequencies were slightly higher (mean values ranged between 25.8° and 36°). As with tones, there were no significant differences between MRAs for narrow-band noise with different center frequencies (a two-way repeated-measures ANOVA with stimulus type and the direction from which the stimuli impinged on the animal as factors indicated a significant effect of direction, $p\leqslant 0.001$, but not of stimulus type and the interaction between stimulus type and direction). Gerbils stimulated from the back performed to 82.2%, 90.0%, 81.7%, 81.7%, 80.6%, and 81.1% correct for frequencies of 0.5, 1, 1.5, 2, 4, and 8 kHz, respectively, at an angle separation of 60°. MRAs for gerbils stimulated from the back were, in general, significantly higher than those of gerbils stimulated from the front [Fig. 4(b); Tukey Test, $P<0.05$]. The only exception was the MRA at 8 kHz ($P=0.403$, Tukey Test). For animals stimulated from the back, the lowest MRA was obtained at 1 kHz; all other MRAs were higher. However, MRAs obtained at the various center frequencies were not significantly different (see ANOVA result).

IV. DISCUSSION

A. Resolution in azimuth sound localization in small mammals

The present study reports accurate sound localization in azimuth in the gerbil for 125-ms pulses of broad-band noise (MRA 22.3°), of narrow-band noise (bandwidth 300 Hz), and of tones of 125 ms duration. The best resolution in azimuth sound localization was found at 2 kHz (22.2° and 20.4° for narrow-band noise and tones, respectively). Below and above this frequency the auditory resolution in azimuth deteriorated. In the only previous psychophysical study of sound localization abilities in the gerbil, Heffner and Heffner (1988a) investigated the resolution in azimuth sound localization for broad-band and tone stimuli using a 2AFC-procedure similar to that employed in the present study. For broad-band noise presented from the front at a fixed level they found a similar, but slightly higher, MRA of 27° than in the present study. The MRA of the gerbil was larger than the 6° of the cat when stimulated from the front (Heffner and Heffner, 1988b). More often, sound localization performance is determined in a relative localization task and reported as the MAA (minimum audible angle). In a MAA task, the stimulus is presented repeatedly and the animal has to report a change in location rather than judge its absolute position as in the MRA task. In general the MAA for broad-band noise varies between 10° and 18° in small mammals (Norway rat: Heffner and Heffner, 1985; least weasel: Heffner and Heffner, 1987; chinchilla: Heffner *et al.*, 1994; old-world fruit bat: Heffner *et al.*, 1999; new-world frugivorous bat: Heffner *et al.*, 2001; big brown bat: Koay *et al.*, 1998), although in some species a much larger MAA was determined (e.g., pocket gopher: Heffner and Heffner, 1990; blind mole rat: Heffner and Heffner, 1992; naked mole rat: Heffner and Heffner, 1993). With the exception of the latter species, stimuli were of such short duration (100–125 ms) that it can be assumed to reflect the open-loop performance of the animals, i.e., the performance in a task that does not allow the animal to adjust its head position in an orientation response to change the binaural cues. Only open-loop experiments enable estimates of the magnitude of the interaural cues that provide for accurate localization.

Heffner and Heffner (1988a) did not determine localization thresholds for tones. Instead, they measured the gerbils' performance (in terms of percent correct responses) in discriminating the location of tones at different frequencies in the azimuthal plane when sounds were presented from one of two speakers separated by an angle of 60°. Only for tones of 32 kHz was the gerbils' discrimination performance well above the threshold criterion of 75% correct responses required for a 2AFC-paradigm. For all other frequencies, gerbils in the study of Heffner and Heffner (1988a) performed

around 75% correct. In the present study, a much higher performance of between 88% and 94% correct responses was found using tones for a separation of 60° between loudspeakers. The MRA for tones in the current study, i.e., the angle at which the gerbils showed 75% correct responses, ranged from 20° at 2 kHz to 30° at 500 Hz. Thus, gerbils in the present study exhibited a better ability to discriminate between speakers at different locations than those in the study of Heffner and Heffner (1988a).

Only relatively few behavioral studies in animals have determined the accuracy of sound-source localization in azimuth for tones or narrow-band noise signals. In the cat, the MRA for tones ranged from 8° at 500 Hz to 26° at 4 kHz (Casseday and Neff, 1973), for a review of localization errors in the cat determined for small angles with head or eye-movements see Populin and Yin, 1998). However, Casseday and Neff repeated the stimuli up to five times which does not exclude the possibility that the cats used closed-loop strategies to locate the sound source. Unfortunately, many studies that sought to determine the MAA for pure tones or narrow-band noise used stimuli of a duration that does not exclude the possibility of closed-loop localization. For tones of 500-ms duration, Martin and Webster (1987) found MAAs ranging from 3.6° at 500 Hz to 25.8° at 32 kHz when determining azimuth sound localization from the front. Brown *et al.* (1978, 1980) determined the MAA for tones and narrow-band noise stimuli of 300 ms duration in macaque monkeys. They determined (Brown *et al.*, 1978) MAAs to range from approximately 3° at 1 kHz to 17° at 11.2 kHz. For tones of 250 Hz the MAA was even larger (22°). The MAA for narrow-band noise with a bandwidth of 250 Hz showed a comparable range of thresholds (5° for a noise centered at 1 kHz up to 17° for a noise centered at 11.2 kHz, Brown *et al.*, 1980). It remains unclear, however, whether the lower level of localization performance of the gerbil was due to procedural differences (e.g., limiting the gerbil to using open-loop strategies only) or some other factor such as smaller available ITD or ILD cues.

B. Localization cues

Azimuth sound localization of tones is dependent on the ability to process ITDs in the low frequency range—the range over which auditory-nerve fibers are able to "phase lock" their action potentials to the stimulus wave form, and on the ability to process ILDs at high frequencies where the head provides a sufficient acoustic shadow (Yin, 2002). Gerbils experience a physiologically relevant range of ITDs of about ±120 μs (Brand *et al.*, 2002). By means of the formula of Kuhn (1977) we calculated the ITD for each threshold angle: $\mathrm{ITD}=(a/c)*s*\sin X$ (a=gerbil head radius of 0.015 m, c=velocity of sound of 343 m/s, X=azimuth angle, s=scale factor that has a value of 3 for frequencies up to 4000 Hz and a value of 2 for 8 kHz, for choice of factor see Kuhn, 1977). In a second step we evaluated the interaural phase differences (IPD =360°*ITD/period). The localization thresholds we measured for tones correspond to ITDs between 40 and 66 μs, and 45 and 111 μs for gerbils stimulated from the front or back, respectively (Fig. 5). For stimulation with narrow-band noise, localization thresholds correspond to ITDs between 44 and 63 μs, and 59 and 103 μs for gerbils stimulated from the front or back, respectively (Fig. 6). These calculated ITDs are similar in magnitude to the ITDs measured by Maki and Furukawa (2005) in the gerbil. Expressed in the form of the equivalent IPD, gerbils show a relative small IPD for stimulation with sound frequencies up to 2 kHz and a strong increase in IPD at the localization thresholds above this frequency. At 8 kHz the IPD approaches 180° for tones and narrow-band noise, respectively, and thus becomes ambiguous (in any case, accurate phase locking is not to be expected at this frequency, although this has not been determined in the gerbil; see Köppl, 1997). It has been demonstrated, however, that transients in the envelopes of signal in which the fine structure of the carrier cannot be resolved can be utilized in the neural computation of ITDs (e.g., Joris, 2003, see also review by Yin, 2002). At frequencies above 4 kHz also ILDs become available for azimuth localization. At the discrimination threshold of about 30° at 8 kHz the gerbil is provided with an ILD of approximately 4 dB (Maki and Furukawa, 2005, and personal communication). That appears to be sufficient for accurate discrimination.

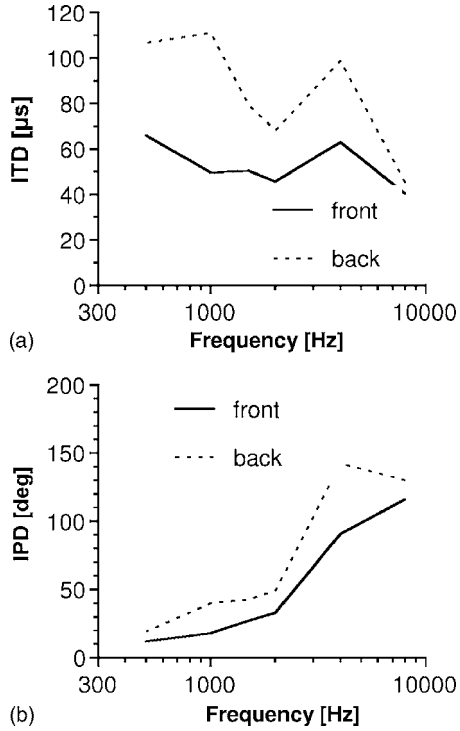


FIG. 5. Interaural time differences [(ITDs), (a)] and interaural phase differences [(IPDs), (b)] corresponding to the mean threshold angles for the localization of tones in relation to tone frequency for stimulation from the front (N=4) and from the back (N=3). ITD was calculated by means of the formula of Kuhn (1977). IPD was derived from the ITD data.

In general the gerbil's MRA was smaller when animals were stimulated from the front compared to when they were stimulated from the back. This difference in auditory resolution in azimuth could be attributed to a number of factors: First, stimulation from the back could provide the subjects with less suitable cues for localization. Measurements of the spatial variation of the gerbil's HRTFs by Maki and Fu-

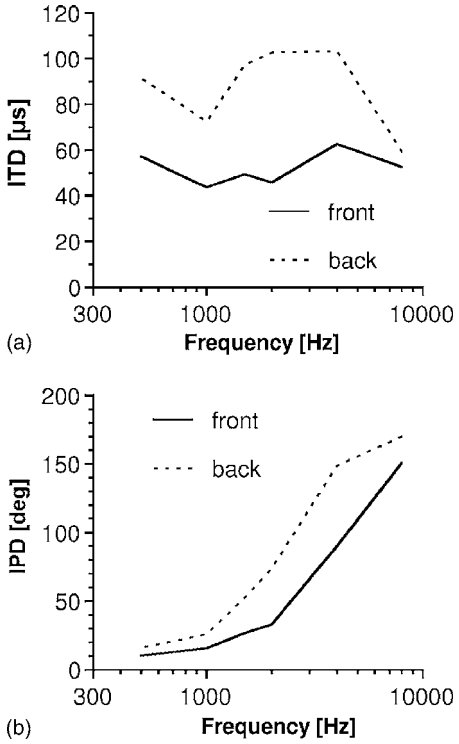


FIG. 6. ITDs (a) and IPDs (b) corresponding to the mean threshold angles for the localization of narrow-band noise in relation to center frequency for stimulation from the front ($N=4$) and from the back ($N=3$). ITD was calculated by means of the formula of Kuhn (1977). IPD was derived from the ITD data.

rukawa (2005, and personal communication; data are available for 5 and 10 kHz among other frequencies) suggest that stimulation from the front provides the gerbil with cues that show a slightly larger monotonic change with a variation of azimuth sound source location than cues available in stimulation from the back. Second, the auditory brain may be specialized to process the direction of sound sources in the frontal hemi-field which includes the field of best vision. Analyzing data on the auditory resolution in azimuth sound localization in a number of animal species, Heffner (1997) showed that the width of the field of best vision is a reliable predictor of the auditory resolution in sound-localization. She argued that this serves to guide orienting movements toward the source of a sound. Thus, accurate localization would only be expected for a range of frontal angles that would elicit exact orienting movements. Finally, when responding to a sound presented from the back turns of greater magnitude must be made, sometimes exceeding 180° when the turn is against the direction from which the sound originated. This may involve more difficult motor patterns and, therefore, result in larger localization errors. Furthermore gerbils, facing away from the speakers, had to store the directional information in their short term memory for a longer time than gerbils facing the speakers. This could have caused additional difficulties for localizing the sound source in the backward stimulation condition.

C. Neural mechanism of sound localization

The superior olivary complex (SOC) is the primary site at which binaural comparisons in the brainstem are made (Yin, 2002). In the SOC the medial superior olive (MSO) and the lateral superior olive (LSO) appear specialized for the neural processing of ITD and ILD cues, respectively. Brand *et al.* (2002) suggested that MSO neurons encode the ITD in the rate of action potentials that changes maximally in the physiologically relevant range of ITDs. As can be deduced from Fig. 2a in Brand *et al.* (2002), the discharge rate of neurons in the gerbil MSO was varied by about 2.6% of the maximum rate for a difference in the ITD of 10 μs. The best ITD calculated from the MRA for tones in the present behavioral study was about 50 μs. This would lead to a reduction in the spike rate by about 13% in neurons of the MSO ipsilateral to the sound source and to an increase of the spike rate by approximately the same amount in neurons of the MSO contralateral to the sound source providing the gerbil's auditory system with a substantial difference in the response at the behaviorally determined threshold for the auditory resolution in azimuth localization. In another rodent with low frequency hearing, the Guinea pig, Shackleton *et al.* (2003) have demonstrated in an analysis of the mean and variance of the response rates of neurons in the inferior colliculus by applying signal-detection theory that single neurons may be able to discriminate ITDs as small as 30 μs. At high frequencies (e.g., at 8 kHz in this study) the change of the response magnitude that occurs in the LSO in neurons in relation to the ILD will be important. Sanes and Rubel (1988) reported a change in the discharge rate of about 2.3% of the maximum rate per dB ILD in LSO neurons of adult gerbils. The estimated ILD of about 4 dB at a MRA of 30° at a frequency of 8 kHz would result in a reduction of the response by about 10% in neurons of the LSO contralateral to the sound source and a similar increase in the response of neurons in the LSO ipsilateral to the sound source. Thus, the relative difference in the response of neurons ipsilateral and contralateral to the sound source is quite similar for the cues providing information at low and high frequencies. Tollin (2004) has demonstrated in the cat that individual neurons in the LSO can detect ILDs as low as 1 dB, although the median of the neuronal ILD discrimination threshold was 4.6 dB. In summary, the data suggest that the changes in the response of neurons in the LSO and MSO in relation to the variation of the sound source azimuth could be sufficiently large to account for the resolution in azimuth sound localization observed in the gerbil in this behavioral study.

ACKNOWLEDGMENTS

We thank Henry and Rickye Heffner as well as Katuhiro Maki and Shigeto Furukawa for providing raw data from their studies. We thank David McAlpine for his comments on an earlier version of the manuscript. This study was supported by the DFG within the international graduate school "Neurosensory Science, Systems and Application" (GRK 591) and the research group "Auditory Objects" (FOR 306).

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Brown, C. H., Beecher, M. D., Moody, D. B., and Stebbins, W. C. (**1978**). "Localization of pure tones by Old World monkeys," J. Acoust. Soc. Am. **63**, 1484–1492.

Brown, C. H., Beecher, M. D., Moody, D. B., and Stebbins, W. C. (**1980**). "Localization of noise bands by Old World monkeys," J. Acoust. Soc. Am. **68**, 127–132.
Carr, C. E., and Konishi, M. (**1988**). "Axonal delay lines for time measurement in the owl's brainstem," Proc. Natl. Acad. Sci. U.S.A. **85**, 8311–8315.
Casseday, J. H., and Neff, W. D. (**1973**). "Localization of pure tones," J. Acoust. Soc. Am. **54**, 365–372.
Heffner, H. E., and Heffner, R. S. (**1985**). "Sound localization in wild Norway rats (*Rattus norvegicus*)," Hear. Res. **19**, 151–155.
Heffner, R. S. (**1997**). "A comparative study of sound localization and its anatomical correlates in mammals," Acta Otorhinolaryngol. Belg. (Stockh) **532**, 46–53.
Heffner, R. S., and Heffner, H. E (**1987**). "Localization of noise, use of binaural cues, and a description of the superior olivary complex in the smallest carnivore, the Least weasel (*Mustela nivalis*)," Behav. Neurosci. **101**, 701–708.
Heffner, R. S., and Heffner, H. E. (**1988a**). "Sound localization and use of binaural cues by the Gerbil (*Meriones unguiculatus*)," Behav. Neurosci. **102**, 422–428.
Heffner, R. S., and Heffner, H. E. (**1988b**). "Sound localization acuity in the cat: Effect of azimuth, signal duration, and test procedure," Hear. Res. **36**, 221–232.
Heffner, R. S., and Heffner, H. E. (**1990**). "Vestigial hearing in a fossorial mammal, the pocket gopher (*Geomys bursarius*)," Hear. Res. **46**, 239–252.
Heffner, R. S., and Heffner, H. E. (**1992**). "Hearing and sound localization in blind mole rats (*Spalax ehrenbergi*)," Hear. Res. **62**, 206–216.
Heffner, R. S., and Heffner, H. E. (**1993**). "Degenerate hearing and sound localization in naked mole rats (*Heterocephalus glaber*) with an overview of central auditory structures," J. Comp. Neurol. **331**, 418–433.
Heffner, R. S., Heffner, H. E., Kearns, D., Vogel, J., and Koay, G. (**1994**). "Sound localization in chinchillas. I. Left/right discriminations," Hear. Res. **80**, 247–357.
Heffner, R. S., Koay, G., and Heffner, H. E. (**1999**). "Sound localization in an old - world Fruit Bat (*Rousettus aegyptiacus*): Acuity, use of binaural cues, and relationship to vision," J. Comp. Psychol. **113**, 297–306.
Heffner, R. S., Koay, G., and Heffner, H. E. (**2001**). "Sound localization in a new-world frugivorous bat, *Artibeus jamaicensis*: Acuity, use of binaural cues, and relationship to vision," J. Acoust. Soc. Am. **109**, 412–421.
Jeffress, L. A. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Joris, P. X. (**2003**). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," J. Neurosci. **23**, 6345–6350.
Klump, G. M., Windt, W., and Curio, E. (**1986**). "The great tit's (Parus major) auditory resolution in azimuth," J. Comp. Physiol. **158**, 383–390.
Koay, G., Kearns, D., Heffner, H. E., and Heffner, R. S. (**1998**). "Passive sound-localization ability of the big brown bat (*Eptesicus fuscus*)," Hear. Res. **119**, 37–48.
Köppl, C. (**1997**). "Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl, *Tyto alba*," J. Neurosci. **17**, 3312–3321.
Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.
Maki, K., and Furukawa, S. (**2005**). "Acoustical cues for sound localization by the Mongolian gerbil Meriones unguiculatus," J. Acoust. Soc. Am. **118**, 872–886; personal communication.
Martin, R. L., and Webster, W. R. (**1987**). "The auditory spatial acuity of the domestic cat in the interaural horizontal and median vertical planes," Hear. Res. **30**, 239–252.
Overholt, E., Rubel, E. W., and Hyson, R. L. (**1992**). "A circuit for coding interaural time differences in the chick brainstem," J. Neurosci. **12**, 1698–1708.
Populin, L. C., and Yin, T. C. T. (**1998**). "Behavioral studies of sound localization in the cat," J. Neurosci. **18**, 2147–2160.
Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.
Ryan, A. (**1976**). "Hearing sensitivity of the mongolian gerbil, *Meriones unguiculatus*," J. Acoust. Soc. Am. **59**, 1222–1226.
Sanes, D. H., and Rubel, E. W. (**1988**). "The ontogeny of inhibition and excitation in the Gerbil lateral superior olive," J. Neurosci. **8**, 682–700.
Shackleton, T. M., Skottun, B. C., Arnott, R. H., and Plmer, A. (**2003**). "Interaural time difference discrimination thresholds for single neurons in the inferior colliculus of Guinea pigs," J. Neurosci. **23**, 716–724.
Tollin, D. J. (2004). "Virtual acoustic space minimum audible angles and interaural level difference discrimination thresholds for single neurons in the lateral superior olive of the cat," Soc. Neuroscience, Prog. No. 304.12.
Yin, T. C. T. (**2002**). "Neural mechanisms of encoding binaural localization cues in the auditory brainstem," in *Integrative Functions in the Mammalian Auditory Pathway*, Springer Handbook of Auditory Research (Springer, New York), pp. 99–159.

Measurement of temporal changes in vocal tract area function from 3D cine-MRI data

Hironori Takemoto and Kiyoshi Honda
ATR Human Information Science Laboratories, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan

Shinobu Masaki, Yasuhiro Shimada, and Ichiro Fujimoto
ATR Brain Activity Imaging Center, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan

(Received 20 December 2004; revised 21 November 2005; accepted 23 November 2005)

A 3D cine-MRI technique was developed based on a synchronized sampling method [Masaki *et al.*, J. Acoust. Soc. Jpn. E **20**, 375–379 (1999)] to measure the temporal changes in the vocal tract area function during a short utterance /aiueo/ in Japanese. A time series of head-neck volumes was obtained after 640 repetitions of the utterance produced by a male speaker, from which area functions were extracted frame-by-frame. A region-based analysis showed that the volumes of the front and back cavities tend to change reciprocally and that the areas near the larynx and posterior edge of the hard palate were almost constant throughout the utterance. The lower four formants were calculated from all the area functions and compared with those of natural speech sounds. The mean absolute percent error between calculated and measured formants among all the frames was 4.5%. The comparison of vocal tract shapes for the five vowels with those from the static MRI method suggested a problem of MRI observation of the vocal tract: data from static MRI tend to result in a deviation from natural vocal tract geometry because of the gravity effect. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151823]

PACS number(s): 43.70.Aj [AL] Pages: 1037–1049

I. INTRODUCTION

Speech sounds result from resonance of a moving vocal tract, and their acoustic properties reflect temporal changes in the vocal tract shape. Studies of vocal tract acoustics have long relied on 2D visualization data obtained by x-ray observation (Chiba and Kajiyama, 1942; Fant, 1960). Recent advancement of digital imaging techniques has led us to visualize 3D shape of the vocal tract during sustainable utterances such as vowels by using magnetic resonance imaging (MRI) (e.g., Rokkaku *et al.*, 1986; Baer *et al.*, 1991) and computed tomography (CT) (e.g., Johansson *et al.*, 1983). Among these techniques, MRI offers a unique means to record dynamic patterns of vocal tract changes including the lower part of the vocal tract during a speech utterance, and it was actually established as a 2D cine-MRI based on a synchronized imaging principle (Masaki *et al.*, 1999). Considering the 3D nature of the changing vocal tract in speech, we further developed a 3D cine-MRI technique to record the time-space pattern of the whole vocal tract including the larynx region during production of a short utterance. This paper reports the method and preliminary results.

MRI has become a technique widely used for speech production studies, because it provides a series of high-resolution tomographic images at any desired orientation without involving harmful effects on subjects (Baer *et al.*, 1991). Since MRI scans usually take more than a few seconds for a single slice and more than a few minutes for multi-slice data with sufficient quality for vocal tract analysis, previous studies have recorded sustainable sounds, such as vowels, fricatives, and liquids (e.g., Alwan *et al.*, 1997; Narayanan *et al.*, 1995, 1997, 1999; Story *et al.*, 1996, 1998, 2001). When multi-slice MRI data are obtained as a volume, the vocal tract area function can be computed by image processing and used to evaluate vocal tract acoustics with transfer function analysis or speech synthesis (Story *et al.*, 1996; Yang and Kasuya, 1994).

While most of the previous studies dealt with still images, attempts have been made to record moving images with a technique called "cine-MRI," which has been used for cardiac motion imaging (e.g., Sechtem *et al.*, 1987). The standard cine-MRI is based on a synchronized sampling principle, which is not real time but similar to a stroboscopic observation of high-speed movements, and thus only applicable to cyclic phenomena such as cardiac movements. More recently, a fast imaging technique called "spiral scan" was adopted in a few studies (e.g., Narayanan *et al.*, 2004). Although such methods suffice to record slow articulatory movements in real time, the image quality and time resolution are generally lower than that of cine-MRI.

A synchronized sampling method with external trigger pulses has been developed by Masaki *et al.* (1999) to record movements of the speech organs as a set of sequential images. Figure 1 shows the experimental setup. The trigger device presents noise burst trains to the subject through a headset and outputs the scan pulses to the MRI scanner to synchronize the data acquisition. The subject listens to the noise burst trains to pace the utterance, while the MRI scanner initiates data acquisition synchronized with the trigger pulses. Stone *et al.* (2001a, b) adopted a similar method using an electrocardiogram (ECG) simulator to control the subject's utterance and MRI scanning for tagged cine-MRI.

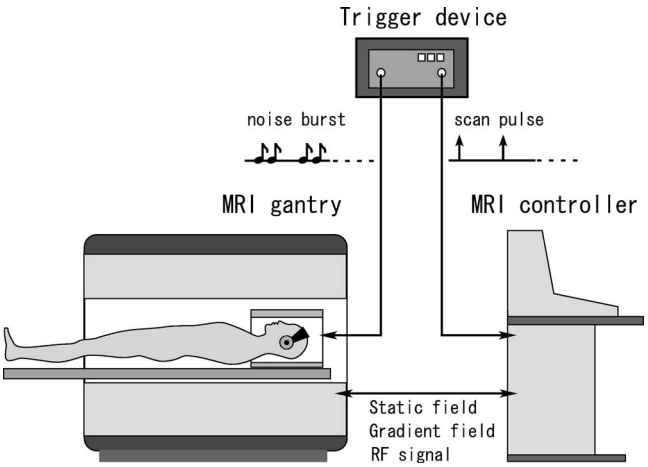


FIG. 1. Experimental setup for the synchronized sampling method. The trigger device, PC-based circuitry, controls the timing of a subject's utterance and that of MRI scanning. The device represents noise burst trains to a subject through the headset, and it sends scan pulses to the MRI controller. The noise burst trains and scan pulses synchronize with each other.

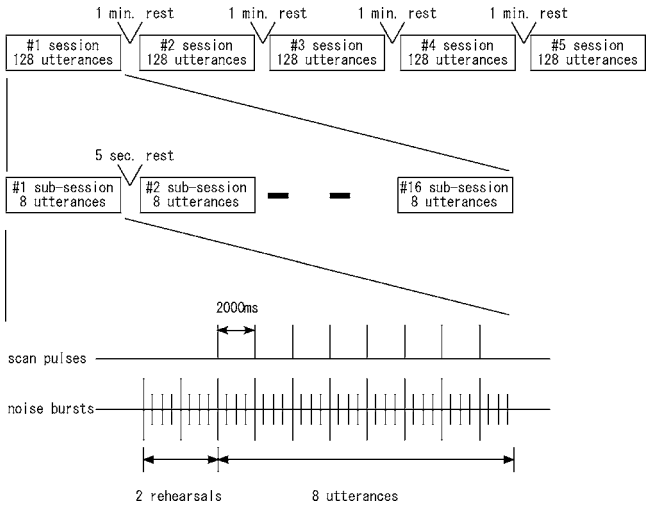


FIG. 2. The subject performed 640 utterances in the 3D cine-MRI experiment. The speech task is divided into five sessions consisting of 16 subsessions, each of which includes eight utterances.

A few different methods have also been attempted. Foldvik *et al.* (1990, 1993, 1995) have recorded articulatory movement using a subject's cardiac R wave as a timing synchronizer. In that method, the R wave's peaks were monitored by ECG and used as the trigger for MRI scanning. Foldvik *et al.* (1993) reconstructed a time series of 3D vocal tract shapes during dynamic articulation from an /a/ to an /i/ configuration, from 5 sagittal and 15 transverse movie slices. Mohammad *et al.* (1997) have recorded articulatory movements without using a timing synchronizer. In their method, after the MRI system acquired data independently of the subject's utterance, *k*-spaces were reorganized off-line by referring to the recorded speech sounds and MRI scan noise during the experiment.

In MRI scanning, magnetic resonance signals are sampled as a two-dimensional matrix in the spatial frequency domain, called *k*-space. Rows in a *k*-space indicate phase encoding steps, while columns represent frequency encoding steps. An MRI scanner repeats data acquisition for rows of the *k*-space one-by-one, gradually changing the phase encoding until the *k*-space is filled. Then, the *k*-space is converted to a real image by the two-dimensional inverse Fourier transform. In common clinical MRI scanners with superconductor magnets, it takes about 10 ms to acquire data for a row. Since the number of rows of a *k*-space is typically 128, it takes about 1280 ms (10 ms$\times$128 steps) to sample an image. In 2D cine-MRI, all the necessary *k*-spaces are prepared prior to the scanning. As the subject performs the *n*th utterance, data for the *n*th rows of all the *k*-spaces are acquired sequentially. When all of the *k*-spaces are filled, they are converted to a set of sequential frames to be observed as a motion picture. The temporal resolution depends on the data acquisition time for filling a row of a *k*-space; if it is 10 ms, the reconstructed data have a rate of 100 frames per second. The spatial resolution is determined by the number of repetitions, and the quality of the reconstructed sequence of images also depends on the subject's consistency in repeated articulations.

The 3D cine-MRI technique is essentially multi-planar 2D cine-MRI. In data acquisition, 2D cine-MRI is repeated in many different slice locations. Since all the images acquired at those planes are time stamped according to the synchronized sampling sequence, they can be arranged to reconstruct a time series of volumetric data. Continuous 3D visualization of the data frame-by-frame permits us to observe articulatory movements as moving 3D images. When the vocal tract region is extracted from each volume, we can directly measure the temporal changes of the area function without relying on the prediction from 2D data.

In the present study, we first describe the methodology of the 3D cine-MRI. Then, we report the result of measurement of the area functions during the vowel sequence /aiueo/. The reliability of the area functions is evaluated by comparing the calculated formant transitions with those of natural speech sounds. In addition, the vocal tract shape of each vowel is compared between dynamic and still conditions to discuss possible problems of dynamic and static MRI methods.

II. MATERIALS AND METHODS

A. Subject and speech task

A 51-years-old male Japanese speaker repeated a continuous utterance of the five Japanese vowels /aiueo/. This utterance is one of the most overlearned phrases among Japanese because it is the first column of the Japanese syllabary and Japanese are trained to recite it in this order like the alphabet in other languages. The vowel /u/ was produced as a mid vowel /ɯ/, as the subject speaks Tokyo dialect. Figure 2 shows the diagram of the task sequence. The utterance was repeated 640 times in total, divided in five sessions of 128 repetitions each. Each session consisted of 16 subsessions, and the subject performed eight utterances for data acquisition after two rehearsals in each subsession. The subject took a rest for 1 min between sessions and for 5 s between subsessions. Figure 2 also indicates the noise burst trains and scan pulses output from a trigger generator program on a

personal computer. The noise burst trains and scan pulses were repeated every 2000 ms. Each cycle included four noise burst signals, and each burst had a 100-ms duration. The sound intensity of the first burst was higher than those of the following three bursts. The subject lay supine in the MRI gantry, wearing a headset for listening to the noise burst signals. During the experiment, the subject breathed in with the mouth closed at the first burst, and began the utterance at the second burst. The total duration of the utterance was approximately 1350 ms.

B. MRI equipment and 3D cine-MRI scan parameters

MRI data were acquired by the Shimadzu-Marconi ECLIPSE 1.5 T PowerDrive 250 installed at the ATR Brain Activity Imaging Center (ATR-BAIC). This MRI scanner is capable of acquiring 2D cine-MRI at an aggregate playback rate of 120 frames per second, and its multi-slice capability allows us to record four parallel planes of cine-MRI simultaneously at a playback rate of 30 frames per second.

The parameters used in the 3D cine-MRI scans were as follows: 3 ms echo time (TE), 1900 ms repetition time (TR), 20 sagittal slice planes, 6 mm slice thickness, 4 mm slice interval, no averaging, 256×256 mm field of view (FOV), and 256×256 pixel image size. The slice interval between neighboring planes was 4 mm. The rightmost plane was located at 39 mm from the midsagittal plane, and the leftmost at 43 mm.

The data thus obtained consist of 20 sagittal sets of 2D cine-MRI with 56 frames for a slice. A tri-linear interpolation algorithm was applied off-line to each set of the sagittal slices so that each volume consists of a cubic element of $1.0\times1.0\times1.0$ mm^3.

C. Choosing vowel center frames in image data

To compare each vowel's area functions from dynamic and static MRI methods for discussing possible problems of these two methods, it is necessary to choose a certain frame from cine-MRI data that represents the center of each vowel. For this purpose, the magnitude of articulatory movement is obtained by computing the mean over a frame of the frame-to-frame absolute difference in each pixel value (mean pixel rate of change, hereafter). When large articulatory movement occurs between neighboring two frames, the mean pixel rate of change increases. In contrast, the value decreases between the frames where articulatory movement is small. Thus, the changes in the mean pixel rate of change correspond to the magnitude of articulatory movements. In the present study, the vowel center is defined as the frame where the mean pixel rate of change becomes a local minimum.

D. Tooth superimposition

MRI has a disadvantage in imaging the teeth because calcified structures lacking mobile hydrogen give no resonance signals, and the region of the teeth is found as dark as the oral cavity. It is therefore necessary to obtain the teeth-air boundary to measure the vocal tract area function accurately from MRI data. We have proposed a method for teeth imaging to solve this problem (Takemoto *et al.*, 2004), which is also applied in the present study. The same subject lay prone in the MRI gantry holding blueberry juice in the mouth as a contrast medium (Hiraishi *et al.*, 1995), while static MRI scan was performed with the following parameters: 11 ms echo time (TE), 3000 ms repetition time (TR), sagittal slice plane, 51 images, 2 mm slice thickness, a 1.5 mm slice interval, no averaging, 256×256 mm field of view (FOV), and 512×512 pixel image size. The images thus obtained demonstrated the oral cavity with high pixel values (bright), while the teeth and jaws appeared with low pixel values (dark). This contrast makes it easy to segment the teeth with the supporting rigid structures from the oral cavity. The maxilla and the mandible with the teeth were reconstructed to obtain the "digital jaw casts," which were then superimposed onto the 56 MRI volumes with the following procedures: (1) The casts were manually superimposed onto the volume data of the first frame; (2) the position of each cast was optimized by rotation and translation along the three axes to minimize the number of voxels in the casts overlapping the surrounding soft tissue of the target volume; (3) the optimized cast position for the first frame was used as the initial cast position for the optimization in the next frame; and (4) the same process was repeated in all the other frames of data. With this procedure, all 56 data volumes were supplemented with the teeth data.

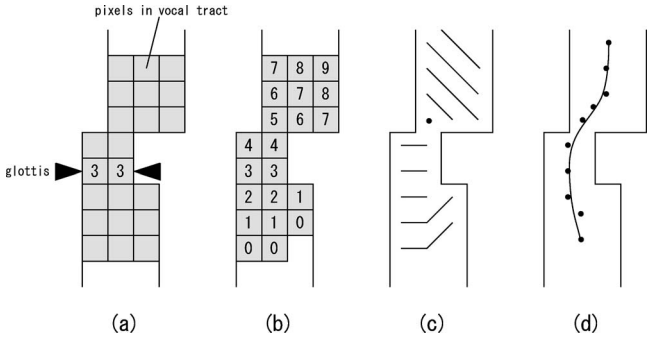


FIG. 3. These figures schematically represent the method for calculating the vocal tract midline in the midsagittal plane near the larynx. (a) Pixels on the glottis are numbered N (in this scheme, $N=3$); (b) pixels in the vocal tract are numbered based on the region growing algorithm from the glottis; (c) lines represent the contour map of numbered pixels; and (d) dots denote the centroids of pixels with the same number, and a spline curve from the centroids is defined as the vocal tract midline.

E. Extracting vocal tract area function

Vocal tract area functions were extracted from the reconstructed volumes with the teeth for the 35 frames from frame 16 (two frames prior to the vowel center for /a/) to frame 50 (one frame subsequent to that for /o/). The extraction was performed frame-by-frame in three steps. First, the vocal tract midline was semi-automatically calculated in the midsagittal image. Then, images perpendicular to the midline were resliced at 2.5-mm intervals along the midline. Finally, the area of the vocal tract region in each section was measured to obtain the area function.

Figure 3 illustrates the custom algorithm developed for calculating the vocal tract midline using a midsagittal model of the larynx. In the first step, the vocal fold line, which

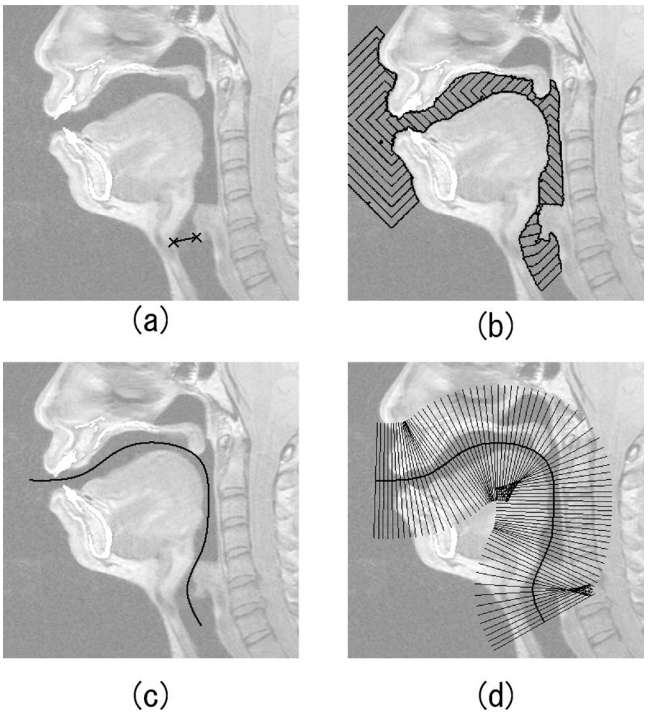


FIG. 4. The algorithm shown in Fig. 3 is applied to an actual image. (a) The vocal fold line is manually selected. (b) A contour map is computed from the numbered pixels. (c) The vocal tract midline is calculated. (d) Lines perpendicular to the midline represent the location of cross sections on which areas are measured.

corresponds to the longitudinal axis of the vocal folds, is manually determined, and an arbitrary positive integer N is assigned to all the pixels along the line [in Fig. 3(a), $N=3$]. In the second step [Fig. 3(b)], a region-growing algorithm assigns an integer number to each pixel lying inside the vocal tract. Note that the pixels superior to the vocal fold line are numbered in an ascending order, while those inferior to the line are numbered in a descending order down to 0. Figure 3(c) shows the contour map of numbered pixels. In the third step, the centroid of each contour is calculated one by one [dots in Fig. 3(d)], and then a spline curve computed from a set of the centroids is defined as the vocal tract midline [curve in Fig. 3(d)]. Figure 4 shows the process of calculating the vocal tract midline on an actual image according to the algorithm described above: Figures 4(a)–4(c) represent the vocal fold line, contour map, and vocal tract midline. Figure 4(d) shows all the locations of cross sections from which the vocal tract area function is measured. The area functions of the bilateral cavities of the piriform fossa were measured in the same cross sections as used for the main vocal tract by pixel counting.

One remaining problem is how to measure the cross-sectional area of the vocal tract near the lip end, where the upper and lower lips are separated and a complete circumferential outline of the vocal tract section cannot be determined. The solution used in this study is to apply the measurable area nearest to the problematic sections. In the furthest section from the glottis where the circumferential area could be measured, the length of the section was extended to half the distance from the end of that section to the last section where the upper and lower lips could still be observed.

F. 3D still MRI data

3D still MRI data were acquired during sustained production of the five Japanese vowels /a/, /i/, /u/, /e/, and /o/, to be compared with the images at the vowel center frames in the 3D cine-MRI data. The same subject took a supine position and produced each vowel maintaining a steady position of the speech organs. The imaging parameters for the data set were as follows: 11 ms echo time (TE), 2700 ms repetition time (TR), sagittal slice plane, 47 images, 2 mm slice thickness, 2 mm slice interval, two times averaging, 256 $\times$256 mm field of view (FOV), and 512$\times$512 pixel image size.

The still image data were processed with the same procedures used for the 3D cine-MRI data to obtain the area functions. In order to compare the position of the speech organs between 3D cine-MRI data and still MRI data, manual tracings were performed to extract the tissue-air boundary and the outline of the hyoid bone in the midsagittal plane.

G. Calculating transfer function

The vocal tract transfer functions were calculated for all the volumes obtained by the dynamic and static MRI using a transmission line model with visco-thermal losses and a yielding wall effect as proposed by Adachi and Yamada (1999). The piriform fossa was implemented as a side branch of the vocal tract in the model using the method proposed by Dang and Honda (1997). The radiation impedance at the open end of the vocal tract was approximated by the method proposed by Causeé *et al.* (1984). The lower four formants were determined from the transfer functions at each frame to be compared with the sound spectra from the subject's natural vowels.

H. Acoustic recording and analyses

Speech sounds were recorded from the subject in a soundproof room. The subject lay supine on the floor with a headset to listen to the noise burst trains and performed the same task of repeating the vowel sequence /aiueo/. The speech signals and noise bursts were recorded with a portable PC-card recorder (MARANTZ, PMD680) at a sampling frequency of 48 kHz. The speech sounds of the five vowels during sustained phonation were also recorded.

From eight repetitions of the vowel sequence /aiueo/, cepstrum-based spectral envelopes were extracted by referring to the noise burst signals to synchronize the analysis window with the image frame. Considering the frame interval in the 3D cine-MRI being 33 ms, the window length and shift length of the analysis are set to 33 ms, and the cepstral order is set at 45. The lower four formants were manually picked from the spectral envelopes frame-by-frame and their mean values over the eight speech samples were obtained. The mean spectral envelope at each frame is also obtained by

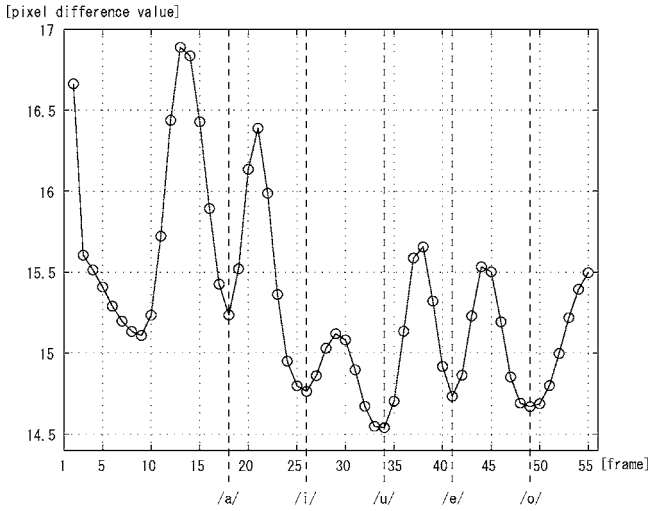


FIG. 5. The time series of the *mean pixel rate of change* are plotted. The local minima except for the first one correspond to the five vowel center frames. The vowel center frames of /a/,/i/,/u/,/e/, and /o/ are 18, 26, 34, 41, and 49, respectively.

averaging eight spectra. The same procedure was used for obtaining the lower four formants for the five vowels recorded during sustained phonation.

III. RESULTS AND DISCUSSION

A. Articulatory movements and vowel center frames in image data

Figure 5 represents the time series of the *mean pixel rate of change* described in Sec. II C. Prephonatory articulatory movements start at frame 10, increase to a local maximum value at frame 14, and then gradually decrease as phonation for /a/ is initiated. At frame 18, the value comes to the first local minimum, which indicates the vowel center for /a/. The other local minima at frames 26, 34, 41, and 49 correspond to the vowel center of /i/, /u/, /e/, and /o/, respectively. The local maxima in the time course of the parameter indicate the magnitude peaks of the articulatory movements. Accordingly, the pattern in Fig. 5 indicates that articulatory movements are the greatest during the transition from /a/ to /i/, while they are the smallest from /i/ to /u/.

B. Visualization of articulatory movements

Figure 6 shows a series of the right half of the orofacial regions of the subject from frame 16 to 50 viewed from the left. The shape and position of the speech organs altered only gradually between neighboring frames. Especially, little change was found among the frames near each of the vowel centers. At all frames, the laryngeal ventricle was observed as a stationary space. The velum begins to rise at the same time as the prephonatory movements start, and it closes the velopharyngeal port completely from frame 18 to 51. In the frames near the center for /a/ (16 to 19) and /o/ (46 to 50), the bilateral palatine tonsils are found to approach the midsagittal plane, thereby decreasing the left-to-right dimension of the pharynx. In those frames, it was also found that the epiglottis makes contact with the tongue root surface, while the apex of the epiglottis is detached from the tongue in the rest of the frames.

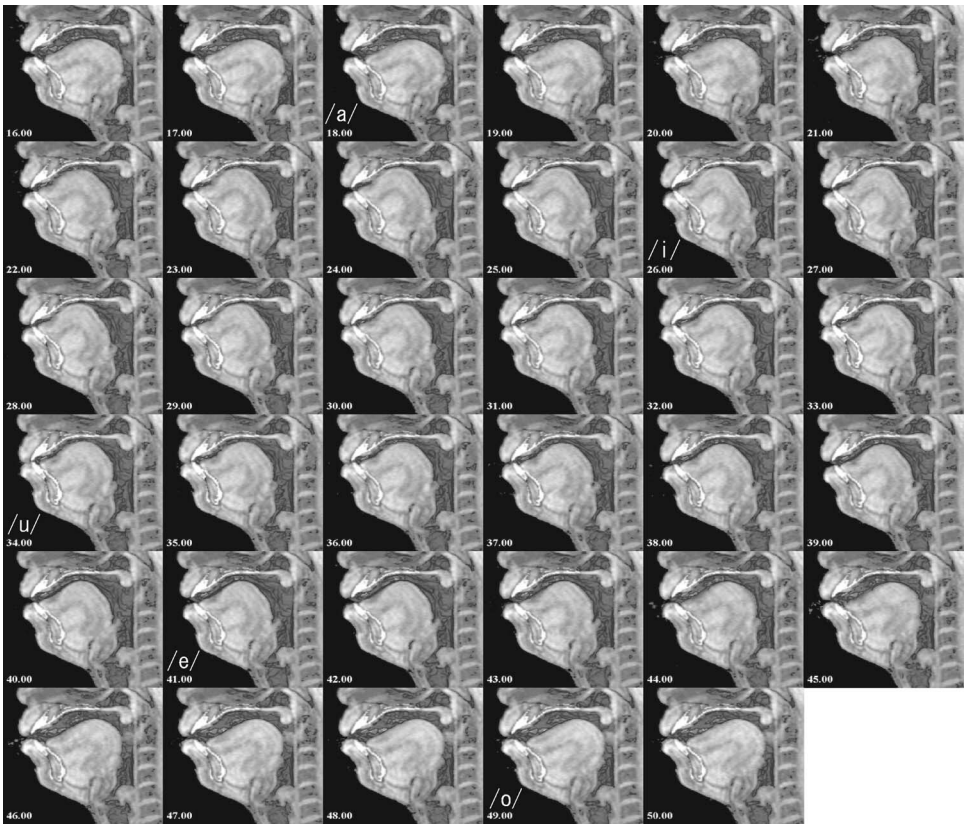


FIG. 6. The right halves of the speech organs were visualized from frame 16 to 50. The upper left is frame 16 and the lower right, frame 50.

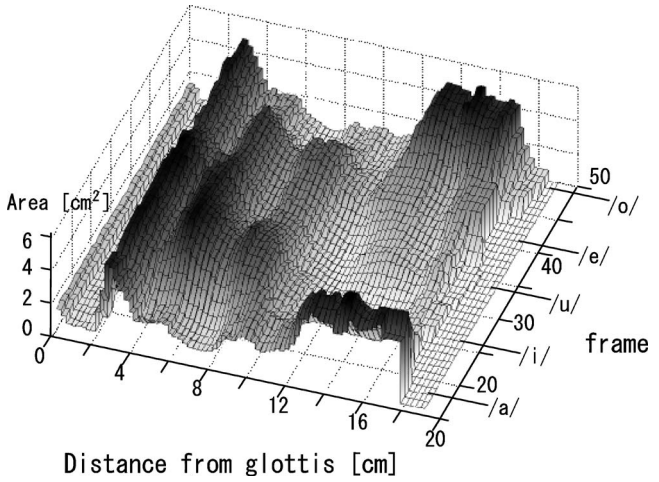


FIG. 7. Temporal changes in area function extracted from 3D cine-MRI data.

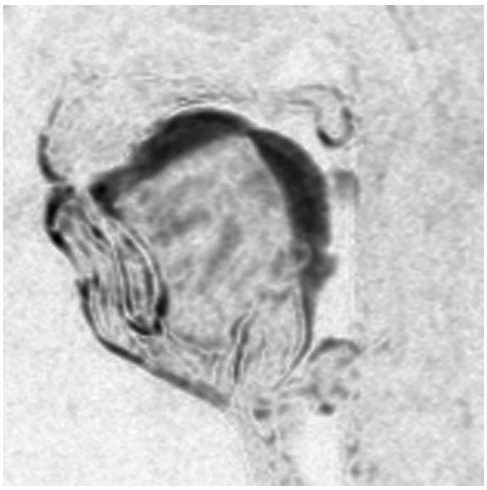


FIG. 8. The standard deviation of each pixel value is calculated in the midsagittal images from frame 16 to 50. The more movements, the darker the pixel. Little movement was observed on the tongue contour near the junction between the soft and hard palates.

C. Temporal pattern of area function

Figure 7 shows the temporal changes of the area function obtained from 3D cine-MRI. The areas in the oral and pharyngeal cavities alter drastically and reciprocally between the front-mid vowels and back vowels. The oral cavity expands and shrinks as a whole, while the pharynx maintains regional details. In the figure, two local minima of area functions are observed in the pharyngeal region throughout the utterance. By referring to the MRI data, they correspond to the constriction formed by the upper free part of the epiglottis around 4 cm from the glottis and the narrowing due to the bilateral palatine tonsils near 8 cm. In contrast to the dynamic changes of these major cavities, there are two smaller regions that are rather stationary throughout the utterance: one is the short segment from the glottis to the wide bottom of the mesopharynx and the other is the point near the junction between the soft and hard palate (about 10 cm from the glottis). The former reflects the structural property of the vocal tract where active articulators are lacking, while the latter is possibly due to functional characteristics of the tongue to regulate vocal tract shape for vowels. Figure 8 shows the standard deviation of each pixel value in the midsagittal images from frame 16 to 50. The dark region in the image indicates the greater regional variation in the image due to articulatory movements. It is clear from the figure that the point on the tongue surface contour opposing the junction of the hard and soft palate maintains a constant location, corresponding to the "pivot" reported by Stone (1990).

D. Temporal changes of three cavity volumes

The descriptions in the previous section regarding Figs. 7 and 8 suggest that the vocal tract can functionally be divided into three portions; front, back, and bottom cavities. The boundary between the front and back cavities is the "pivot" (10.75 cm from the glottis), and the bottom cavity corresponds to the laryngeal ventricle and vestibular space bounded by the glottis and the hypopharyngeal widening at the peak in the area function (3 cm from the glottis).

To characterize the temporal changes of the three cavities, Fig. 9 shows the time course of their volumes. The volume of the bottom cavity is almost constant during speech, with the standard deviation of 0.39 cm^3. In contrast, the back and front cavities show a marked change in volume, with a greater standard variation in the front cavity (8.80 cm^3) than in the back cavity (5.36 cm^3). The largest change is observed in both cavities from /a/ to /i/: the front cavity volume decreases to less than one-sixth and the back cavity volume increases to more than twice. From /i/ to /u/, the change in the front cavity volume is smaller than that of the back cavity volume, although the front cavity volume doubles. From /e/ to /o/, the front cavity increases while the back cavity decreases in volume, but the magnitude of these changes is smaller than that between /a/ and /i/. These observations suggest that the front and back cavity volumes change in a reciprocal manner.

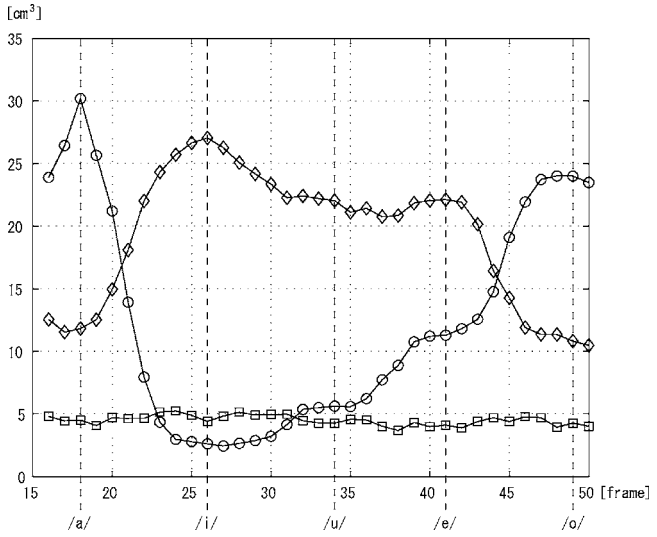


FIG. 9. Temporal changes of three cavity volumes. The volume change of the bottom cavity is depicted by the square symbols, that of the back cavity by the diamonds, and that of the front cavities by the circles.

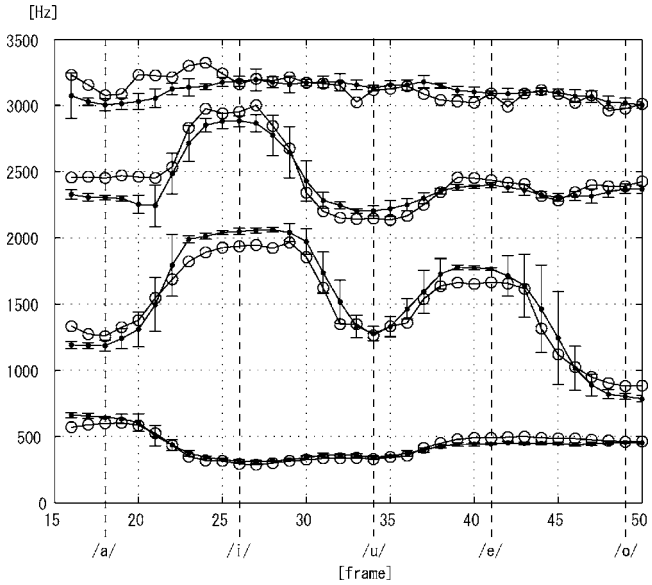


FIG. 10. Calculated and measured transition pattern of the first four formants. Circles indicate the temporal pattern of the lower four formants of the transfer function. Dots with error bars (one standard deviation) indicate the mean formant values from eight speech samples.

The mean pixel rate of change shown earlier in Fig. 5 approximately reflects the changes in volumes of the front and back cavities. The intervowel frames where a large volume change is observed indicate the larger value of the mean pixel rate of change (e.g., from /a/ to /i/). In contrast, the frames where small volume change occurs also exhibit smaller changes in the mean pixel rate of change (from /i/ to /u/). Although the change in volume of the back cavity is small from /u/ to /e/, the mean pixel rate of change increases due to the jaw opening gesture.

E. Acoustic evaluation of area functions from 3D cine-MRI data

A question can be raised as to whether the area functions obtained from the 3D cine-MRI data correspond to those during natural condition. In order to evaluate the reliability of the area functions extracted from the 3D cine-MRI data, the temporal pattern of the lower four formants was calculated and compared with that of natural speech sounds. Figure 10 shows calculated and measured transition patterns. Circles indicate the formant frequencies measured from the transfer functions, and dots with error bars (one standard deviation) show the mean formant values obtained from eight speech samples.

The mean absolute percent error between all the formant data from the transfer functions and natural speech sounds is 4.5%. Relatively larger errors are found regionally: from frame 16 to 20 (around /a/), the third formant of the transfer function is higher than that of natural speech; from frame 23 to 28 (around /i/) and from 39 to 41 (around /e/), the calculated second formant is lower than the measured one.

The standard deviations of the formants in natural speech (shown by the error bars) are larger at the intervowel frames than at the vowel center frames. This contrast indicates that during repetitions of the utterance, articulatory movements fluctuate from token to token at the transitional frames, while they are relatively invariable at the vowel center frames.

Detailed acoustic evaluation was performed for the five vowels in 3D cine-MRI data. Tables I and II list numerical data for the area functions of the main vocal tract at the five vowel center frames, including the midsagittal dimensions in parentheses for the purpose of reference. Table III gives the area functions of the right and left cavities of the piriform fossa. Table IV shows the lower four formants obtained from the transfer functions, those measured from the natural speech, and the percent errors for the former relative to the latter, at the five vowel center frames.

Figure 11 shows the simulated transfer functions and natural speech spectra at the centers of the five vowels to illustrate the overall spectral resemblance and local differences. The mean absolute percent error between the measured and calculated lower four formants was 3.9%. Although similarities are found among the lower four formants, differences are noted in the higher frequency regions above 3.1 kHz. In all the vowels, the calculated fifth formant tends to be positioned higher than the measured one. Between 4 to 5 kHz, the bilateral cavities of the piriform fossa generate two zeros on the transfer function of vowel spectra, consistent with the previous study (Dang and Honda, 1997). The locations of zeros in the transfer functions roughly coincided with those of minima on the speech spectra in each vowel.

F. Acoustic evaluation of area functions from static MRI data

Before comparing the vocal tract shapes between dynamic and static conditions, an acoustic evaluation is performed for the area functions extracted from the still MRI data. Tables V and VI list numerical data for the area functions of the main vocal tract for the five vowels, and Table VII shows those of the left and right cavities of the piriform fossa. Table VIII indicates the lower four formants obtained from the transfer functions, those from natural speech, and the percent errors. Although the mean absolute percent error between the calculated and measured formants was 4.9%, a large error is observed in the second formant of the vowel /a/ (28.8%).

G. Comparison of vocal tract shape for five vowels between dynamic and static conditions

It has often been observed that midsagittal articulatory configurations for vowels differ slightly between dynamic and still MRI data (Honda *et al.*, 2004), and differences are observed in the tongue tip location, tongue body shape, and jaw position. In this section, we compare the vocal tract shapes obtained under the static and dynamic MRI, and point out possible factors causing deviation from natural articulation in static MRI.

Figure 12 shows the midsagittal traces of the speech organ (left column) and area functions (right column) of the five vowels in the 3D cine-MRI data and still MRI data. These tracings show that under the static condition, the tongue body tends to be more advanced and higher in the

TABLE I. Main vocal tract area functions from sections 1–39 for the five vowels extracted from the 3D cine-MRI data. Units are cm^2. Each section length is 0.25 cm. The left and right piriform fossae connected at the 11th section in all the vowels. The value in the parentheses indicates the midsagittal dimension of the section (cm).

Section no.	/a/	/i/	/u/	/e/	/o/
1	1.45 (1.20)	1.28 (1.10)	1.18 (0.50)	1.41 (1.10)	1.43 (1.00)
2	1.09 (0.90)	1.04 (0.90)	1.05 (1.30)	1.18 (1.00)	1.23 (1.40)
3	0.48 (1.10)	0.51 (1.10)	0.51 (0.80)	0.57 (1.00)	0.63 (1.10)
4	0.44 (0.50)	0.46 (0.60)	0.48 (0.80)	0.48 (0.50)	0.51 (0.70)
5	0.50 (0.50)	0.46 (0.50)	0.51 (0.70)	0.46 (0.40)	0.53 (0.60)
6	0.48 (0.30)	0.48 (0.40)	0.51 (0.50)	0.48 (0.30)	0.53 (0.40)
7	0.54 (0.30)	0.48 (0.40)	0.53 (0.40)	0.51 (0.30)	0.53 (0.30)
8	0.48 (0.20)	0.46 (0.30)	0.55 (0.30)	0.52 (0.30)	0.59 (0.30)
9	0.60 (0.30)	0.57 (0.30)	0.53 (0.40)	0.50 (0.30)	0.57 (0.30)
10	3.13 (0.60)	2.25 (0.50)	1.84 (0.60)	1.85 (0.60)	1.65 (0.50)
11	4.59 (1.50)	4.72 (1.80)	4.20 (1.30)	4.13 (1.50)	4.19 (1.10)
12	4.24 (1.40)	4.89 (1.70)	5.16 (1.80)	4.31 (1.60)	4.71 (1.60)
13	3.61 (1.40)	4.79 (1.70)	4.91 (1.80)	3.89 (1.60)	4.18 (1.60)
14	3.20 (1.30)	4.24 (1.80)	4.77 (1.90)	3.36 (1.50)	3.80 (1.60)
15	2.67 (1.20)	3.75 (1.70)	4.37 (1.80)	3.18 (1.50)	3.24 (1.50)
16	1.87 (1.00)	3.47 (1.60)	4.15 (1.80)	2.61 (1.30)	2.66 (1.30)
17	1.00 (0.90)	3.00 (1.50)	3.52 (1.70)	2.13 (1.20)	2.11 (1.10)
18	1.27 (0.60)	3.30 (1.60)	3.38 (1.60)	2.66 (1.30)	1.71 (0.90)
19	1.14 (0.60)	4.35 (1.50)	4.20 (1.40)	3.01 (1.00)	2.10 (0.90)
20	1.58 (0.50)	4.34 (1.40)	3.92 (1.40)	2.98 (1.10)	1.86 (0.70)
21	1.48 (0.60)	4.71 (1.40)	4.17 (1.40)	3.73 (1.50)	2.10 (0.90)
22	1.61 (0.90)	4.94 (1.70)	4.38 (1.50)	4.04 (1.70)	1.99 (1.00)
23	1.60 (0.90)	4.67 (2.00)	4.17 (1.80)	3.77 (1.80)	1.80 (1.00)
24	1.18 (0.70)	4.47 (2.00)	3.81 (1.80)	3.65 (1.90)	1.38 (0.90)
25	0.99 (0.70)	4.31 (2.10)	3.29 (1.80)	3.37 (1.90)	1.03 (0.80)
26	0.89 (0.70)	3.96 (2.20)	2.90 (1.80)	2.94 (2.00)	0.84 (0.70)
27	0.76 (0.70)	3.50 (2.20)	2.54 (1.90)	2.68 (2.10)	0.66 (0.70)
28	0.64 (0.70)	3.46 (2.30)	2.13 (1.80)	2.47 (2.10)	0.42 (0.60)
29	0.64 (0.80)	3.28 (2.40)	1.90 (1.80)	2.28 (2.20)	0.40 (0.40)
30	0.68 (0.90)	3.10 (2.60)	1.84 (1.90)	2.31 (2.30)	0.36 (0.30)
31	0.76 (0.80)	3.46 (2.70)	1.70 (1.80)	2.63 (2.50)	0.34 (0.30)
32	0.83 (0.70)	3.69 (2.80)	2.03 (2.10)	2.96 (2.60)	0.34 (0.40)
33	0.89 (1.50)	4.18 (2.90)	2.46 (2.20)	3.21 (2.10)	0.40 (0.40)
34	1.28 (1.60)	3.58 (2.10)	2.78 (2.00)	3.09 (1.90)	0.76 (1.20)
35	1.63 (1.10)	3.48 (2.00)	2.53 (1.50)	3.12 (1.80)	1.21 (1.50)
36	1.77 (0.90)	3.56 (2.20)	2.49 (1.50)	3.27 (2.00)	0.93 (0.60)
37	1.84 (1.00)	3.23 (2.10)	2.37 (1.50)	3.00 (1.90)	0.80 (0.50)
38	2.06 (1.20)	2.98 (2.00)	2.08 (1.60)	2.84 (1.90)	0.99 (0.50)
39	2.34 (1.50)	2.32 (1.60)	1.71 (1.30)	2.40 (1.70)	1.14 (0.90)

vowels /i/ and /u/, more retracted in the vowels /a/ and /o/, and higher in the vowel /e/. In the static MRI data, the tongue body appears more spherical in the non-close vowels (/e/, /a/, and /o/). In the close vowels (/i/ and /u/), the constriction in the oral cavity is narrower in the static condition than in the dynamic condition. The area functions show that in the static condition the oral cavity volume is smaller in all the vowels, partly reflecting the more closed jaw positions. Also, the areas near the soft palate (about 10 cm from the glottis) are smaller in /a/ and /o/, and the pharyngeal cavity volume is generally slightly smaller under the static condition, while it is exclusively larger in /i/. The areas in the larynx for the five vowels are smaller under the static condition.

These differences noted above may be attributed to various factors. First, the tongue would be influenced by the accumulated effect of gravity during sustained vowel production, thereby tending to reduce the pharyngeal volume in open vowels. To resist against gravity during static MRI experiments, subjects would use the hard palate as a support for the tongue to maintain its high position, which could result in an extreme articulation with a smaller oral volume in close vowels. Further, the gravity force acts on the mandible, which would limit forward translation of the condyle in producing open vowels and result in a smaller oral cavity volume. Second, subjects in static MRI experiments tend to save breath by narrowing the laryngeal vestibule to maintain the same articulatory posture with minimum respiratory movement. This effort would result in a smaller volume in the laryngeal region. Third, the loud machine noise generated from the MRI scanner during the scans prevents subjects' auditory monitoring of vowel production, which would alter

TABLE II. Main vocal tract area functions from section 40 to lip end. The bottom line of the table shows the total length of the vocal tract.

Section no.	/a/	/i/	/u/	/e/	/o/
40	2.18 (1.50)	1.92 (1.30)	1.29 (1.10)	1.88 (1.20)	1.16 (1.00)
41	1.88 (1.40)	1.55 (1.10)	0.92 (0.70)	1.81 (1.00)	1.00 (1.00)
42	1.52 (1.10)	1.32 (0.80)	0.80 (0.80)	1.58 (1.10)	0.88 (0.90)
43	1.49 (0.90)	1.26 (1.00)	0.64 (0.60)	1.61 (1.00)	0.69 (0.70)
44	1.58 (0.90)	0.80 (0.50)	0.61 (0.70)	1.30 (0.90)	0.79 (0.80)
45	1.69 (1.10)	0.57 (0.40)	0.45 (0.50)	1.18 (0.80)	0.87 (0.70)
46	2.24 (1.10)	0.46 (0.30)	0.31 (0.40)	1.22 (0.70)	1.02 (0.80)
47	2.50 (1.10)	0.38 (0.10)	0.29 (0.30)	1.48 (0.70)	1.09 (0.70)
48	2.98 (1.10)	0.31 (0.10)	0.37 (0.30)	1.68 (0.70)	1.35 (0.80)
49	3.71 (1.20)	0.25 (0.20)	0.39 (0.30)	1.75 (0.60)	1.79 (0.80)
50	4.29 (1.30)	0.26 (0.10)	0.43 (0.30)	1.86 (0.70)	2.30 (1.00)
51	4.61 (1.30)	0.28 (0.10)	0.83 (0.40)	1.92 (0.80)	2.92 (1.10)
52	4.89 (1.40)	0.24 (0.10)	1.09 (0.40)	1.81 (0.80)	4.01 (1.30)
53	5.08 (1.40)	0.20 (0.10)	1.12 (0.40)	1.95 (0.80)	4.20 (1.40)
54	5.22 (1.40)	0.11 (0.10)	1.14 (0.50)	1.91 (0.70)	4.65 (1.40)
55	5.32 (1.50)	0.17 (0.20)	1.16 (0.50)	1.96 (0.60)	4.67 (1.50)
56	5.38 (1.50)	0.17 (0.20)	1.29 (0.40)	1.88 (0.70)	4.51 (1.40)
57	5.36 (1.50)	0.21 (0.10)	1.27 (0.40)	1.56 (0.70)	4.56 (1.40)
58	5.34 (1.40)	0.21 (0.20)	1.30 (0.50)	1.43 (0.80)	4.61 (1.50)
59	5.24 (1.30)	0.23 (0.10)	1.02 (0.40)	1.31 (0.80)	4.01 (1.30)
60	5.15 (1.40)	0.25 (0.30)	1.06 (0.60)	1.21 (0.80)	4.43 (1.30)
61	4.83 (1.40)	0.31 (0.30)	0.88 (0.60)	1.27 (0.90)	5.19 (1.40)
62	4.48 (1.30)	0.31 (0.50)	0.85 (0.90)	1.36 (0.80)	4.59 (1.50)
63	4.21 (1.10)	0.38 (0.40)	0.58 (0.60)	1.96 (0.80)	4.67 (1.40)
64	4.05 (1.00)	0.35 (0.50)	0.67 (0.50)	1.93 (0.70)	4.26 (1.20)
65	4.39 (1.10)	0.66 (0.50)	1.07 (0.50)	2.29 (0.70)	4.20 (1.10)
66	4.63 (1.20)	0.82 (0.40)	1.48 (0.60)	2.23 (0.70)	4.30 (1.10)
67	4.72 (1.40)	0.86 (0.40)	1.41 (0.60)	2.23 (0.80)	4.56 (1.20)
68	4.72 (1.30)	0.55 (0.30)	0.75 (0.40)	2.23 (0.80)	4.38 (1.50)
69	4.72 (0.80)	0.55 (1.10)	0.16 (0.30)	2.23 (1.00)	2.93 (1.10)
70	4.72 (0.60)	0.55 (1.50)	0.16 (0.30)		2.05 (0.90)
71	4.72 (0.70)		0.16 (0.30)		1.15 (0.50)
72			0.16 (0.20)		0.67 (0.40)
73					0.63 (0.40)
74					0.63 (0.40)
VT length	17.750	17.500	18.000	17.250	18.500

articulatory status from a natural one to a sustainable setting for the situation with poor auditory feedback. Although the machine noise could also affect subjects' articulation during dynamic MRI scans, its effect could be smaller in the case of the familiar utterance used in this study.

IV. CONCLUSION

The present study described the 3D cine-MRI method, reported the changes in the area functions during a continuous utterance /aiueo/ in Japanese, and performed acoustic

TABLE III. Equal interval (0.25 cm) area functions of the piriform fossa for the five vowels extracted from the 3D cine-MRI data. The bottom line of the table indicates the total length of the piriform fossa.

	/a/		/i/		/u/		/e/		/o/	
Section no.	left	right	left	right	left	right	left	right	left	right
1	0.84	0.75	0.9	0.77	0.80	0.84	0.90	0.74	0.99	0.82
2	0.72	0.69	0.84	0.68	0.75	0.73	0.72	0.66	0.83	0.74
3	0.67	0.62	0.63	0.70	0.70	0.64	0.68	0.57	0.67	0.69
4	0.55	0.52	0.68	0.45	0.80	0.46	0.59	0.64	0.71	0.68
5	0.26	0.35	0.57	0.25	0.52	0.39	0.37	0.40	0.44	0.52
6	0.10	0.21	0.37	0.15	0.31	0.14	0.12	0.21	0.21	0.30
7					0.05					0.22
Total length	1.50	1.50	1.50	1.50	1.75	1.50	1.50	1.50	1.50	1.75

TABLE IV. Lower four formants of transfer functions at the five vowel center frames from the 3D cine-MRI data, those of natural speech, and percent errors for the former relative to the latter. "t" denotes the transfer function, "n" the natural speech, and "d" percent errors.

	/a/	/i/	/u/	/e/	/o/
tF1	600	295	330	490	460
tF2	1260	1935	1265	1665	880
tF3	2450	2950	2145	2435	2390
tF4	3075	3160	3115	3090	2975
nF1	644	314	343	445	456
nF2	1188	2047	1279	1763	803
nF3	2303	2882	2203	2399	2370
nF4	3004	3181	3129	3091	3016
dF1	−6.8	−6.1	−3.8	10.1	0.9
dF2	6.1	−5.5	−1.1	−5.6	9.6
dF3	6.4	2.4	−2.6	1.5	0.8
dF4	2.4	−0.7	−0.4	0.0	−1.4

evaluation of the area functions. The mean absolute percent error (4.5%) between all the formant data from the transfer functions and natural speech indicates that the changes in the area functions are successfully measured by the 3D cine-MRI technique.

The volume of the bottom cavity (from the glottis to the lower pharynx) was almost constant during the utterance (0.39 cm^3 standard deviation), which is consistent with the MRI analysis by Kitamura *et al.* (2005a) for hypopharyngeal cavity geometry. The vocal tract area near the junction between the soft and hard palate demonstrated the smallest area changes in the vocal tract (Fig. 8), corresponding to the "pivot" reported by Stone (1990). The volumes of the back cavity (from the lower pharynx to the "pivot") and front cavity (from the "pivot" to the lip ends) indicated the tendency of reciprocal changes (Fig. 9). This is consistent with the general characteristic observed in natural speech. The previous statistical analyses based on 2D cine-radiography

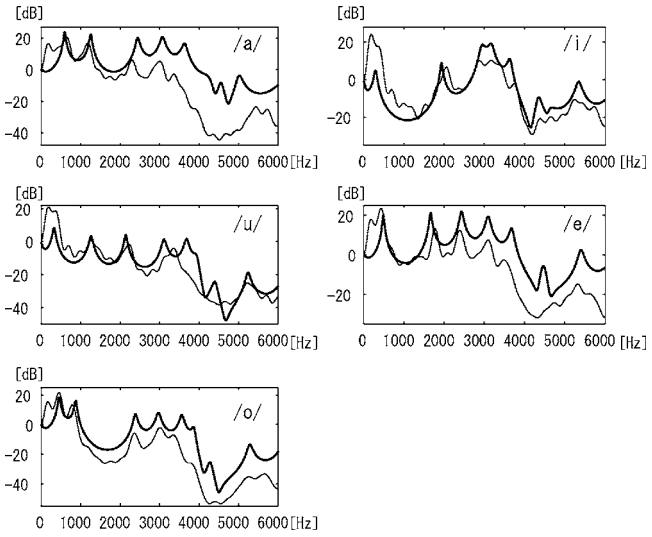


FIG. 11. The transfer function and cepstrum-based speech spectrum are plotted together at the five vowel center frames. The bold line represents the transfer function, and the thin line denotes the speech spectrum.

TABLE V. Main vocal tract area functions from sections 1–40 for the five vowels extracted from the static MRI data. Units are cm^2. Each section length is 0.25 cm. "(l)" denotes the section connected to the left piriform fossa, "(r)" denotes the section connected to the right piriform fossa, and "(l, r)" denotes the section connected to the left and right piriform fossae.

Section no.	/a/	/i/	/u/	/e/	/o/
1	0.468	0.782	0.782	0.615	0.590
2	0.425	0.664	0.684	0.540	0.500
3	0.360	0.430	0.490	0.417	0.430
4	0.332	0.391	0.468	0.370	0.370
5	0.313	0.332	0.332	0.325	0.352
6	0.254	0.254	0.274	0.253	0.313
7	0.215	0.313	0.332	0.215	0.293
8	0.215	0.508	0.405	0.333	0.332
9	0.332	0.877	0.575	0.542	0.391
10	2.105(l)	4.440	1.212	1.802(l)	1.965(l)
11	3.873(r)	5.232(l, r)	4.433(l, r)	3.908(r)	3.780(r)
12	3.803	5.322	4.678	3.962	3.877
13	3.490	5.385	4.555	3.763	3.603
14	3.083	5.317	4.503	3.535	3.297
15	2.817	5.272	4.240	3.285	2.947
16	2.235	5.045	3.953	2.993	2.438
17	1.745	4.987	3.647	2.690	1.940
18	1.712	5.270	3.857	2.717	1.850
19	1.800	5.706	4.082	3.013	2.143
20	1.885	5.725	4.000	3.132	2.105
21	1.833	5.667	4.130	3.275	2.067
22	1.835	5.491	3.913	3.170	2.147
23	1.732	5.354	3.632	3.197	1.962
24	1.410	5.159	3.325	3.060	1.500
25	1.212	4.866	3.030	2.925	1.198
26	1.058	4.533	2.765	2.700	0.960
27	0.983	4.348	2.397	2.498	0.787
28	0.925	4.200	2.292	2.348	0.660
29	0.858	4.207	2.100	2.320	0.547
30	0.820	4.372	2.070	2.380	0.422
31	0.890	4.705	2.280	2.817	0.435
32	0.935	4.808	2.837	3.045	0.435
33	1.175	4.572	2.940	2.286	0.490
34	1.368	4.084	2.443	2.285	0.718
35	1.211	3.615	2.247	2.105	1.045
36	0.801	3.400	2.025	2.005	1.094
37	0.762	2.951	1.990	1.823	0.313
38	0.879	2.579	1.718	1.450	0.258
39	0.957	2.032	1.340	1.125	0.158
40	1.290	1.446	1.002	0.833	0.212

(e.g., Shirai and Honda, 1976; Harshman *et al.*, 1977; Maeda, 1990) have equally shown that the front-back reciprocity constitutes the first principal component.

The comparison of the vocal tract shapes under the dynamic and static conditions suggested that static MRI data involve effects of gravity force, phonatory effort, and machine noise. The gravity effect during MRI experiment has been examined using an open-type MRI scanner (Kitamura *et al.*, 2005b), which partly agrees with our observation in this study. Although the gravity effect is unavoidable as long as subjects lie supine during MRI experiments, it could be minimized with an intermittent scan sequence (Story *et al.*, 1996) that allows subjects to reset the long-term effect. Also, the technique of synchronized cine-MRI as described in this

TABLE VI. Main vocal tract area functions from section 41 to lip end for five vowels extracted from the static MRI data. Each section length is 0.25 cm. The bottom line of the table displays the total length of the vocal tract.

Section no.	/a/	/i/	/u/	/e/	/o/
41	1.133	1.036	0.625	0.618	0.542
42	0.977	0.801	0.332	0.535	0.570
43	0.879	0.645	0.195	0.477	0.453
44	0.897	0.332	0.117	0.468	0.410
45	0.953	0.215	0.098	0.630	0.410
46	1.218	0.137	0.190	0.750	0.449
47	1.593	0.078	0.268	0.830	0.528
48	2.185	0.137	0.323	0.860	0.688
49	2.735	0.234	0.449	0.973	1.115
50	3.337	0.274	0.547	0.970	1.720
51	3.737	0.234	0.625	1.030	2.510
52	3.965	0.212	0.664	1.115	2.985
53	4.218	0.185	0.698	1.045	3.410
54	4.372	0.130	0.695	0.963	3.763
55	4.555	0.113	0.623	0.860	3.985
56	4.607	0.072	0.593	0.802	4.160
57	4.515	0.058	0.532	0.662	4.345
58	4.372	0.059	0.407	0.705	4.537
59	4.100	0.098	0.352	0.667	4.725
60	3.763	0.090	0.290	0.570	4.470
61	4.660	0.195	0.230	0.427	4.383
62	4.385	0.234	0.147	0.367	4.247
63	4.263	0.371	0.137	0.383	4.360
64	3.880	0.528	0.098	0.660	4.195
65	3.922	0.860	0.176	1.373	4.048
66	4.303	1.397	0.455	1.373	3.820
67	4.465	1.397	1.062	1.373	3.973
68	4.793		1.205	1.373	4.245
69	5.002		0.957		4.402
70	5.002		0.375		3.938
71	5.002		0.215		3.220
72	5.002				2.250
73					1.407
74					0.642
75					0.254
VT length	18.000	16.750	17.750	17.000	18.750

TABLE VIII. Lower four formants of vocal tract transfer functions from the static MRI data, those of natural speech, and percent errors for the former relative to the latter. "t" denotes the transfer function, "n" the natural speech, and "d" percent errors.

	/a/	/i/	/u/	/e/	/o/
tF1	590	250	305	410	440
tF2	1385	2145	1565	1790	805
tF3	2505	3110	2285	2445	2495
tF4	3305	3365	3330	3325	3230
nF1	581	272	325	382	414
nF2	1078	2041	1517	1752	772
nF3	2629	3004	2242	2454	2464
nF4	3192	3303	3190	3270	3187
dF1	1.5	−8.1	−6.2	7.3	6.3
dF2	28.5	5.1	3.2	2.2	4.3
dF3	−4.7	3.5	1.9	−0.4	1.3
dF4	3.5	1.9	4.4	1.7	1.3

study offers a similar approach. Other problems associated with static MRI such as phonatory effort and machine noise could relatively easily be handled by some adequate means.

The advantage of the 3D cine-MRI method is the capability of measuring temporal change of the whole vocal tract during a continuous utterance. This is virtually a unique method for obtaining 3D movement of the vocal tract including the lower pharyngeal region. A possible application of this method is measuring movements of the lower pharynx and larynx during changes in voice quality and fundamental frequency. Another is to measure the changes in the shape and size of the interdental space, whose side-branch geometry would generate spectral zeros.

The disadvantage of the 3Dcine-MRI method is the necessity for many repetitions of utterances (640 times in the present study). The large number of repetitions would increase chances of motion artifacts and could cause irrecoverable error due to gradual changes in head orientation. These problems may be solved in future by high-speed volume imaging techniques such as the parallel image acquisition with a custom array coil (Pruessmann *et al.*, 1999).

TABLE VII. Equal interval (0.25 cm) area functions of piriform fossae for the five vowels extracted from the static MRI data. The bottom line of the table indicates the total length of the piriform fossa.

	/a/		/i/		/u/		/e/		/o/	
Section no.	left	right	left	right	left	right	left	right	left	right
1	0.680	0.650	0.970	0.860	0.980	0.700	0.530	0.650	0.620	0.450
2	0.560	0.490	0.680	0.600	0.690	0.500	0.460	0.560	0.550	0.380
3	0.510	0.410	0.600	0.550	0.560	0.470	0.450	0.480	0.460	0.300
4	0.480	0.280	0.500	0.460	0.560	0.370	0.370	0.360	0.390	0.230
5	0.330	0.150	0.470	0.230	0.520	0.230	0.300	0.250	0.320	0.100
6	0.220	0.040	0.400	0.160	0.410	0.120	0.190	0.100	0.200	0.030
7	0.040		0.150	0.110	0.340	0.010	0.090	0.010	0.020	
8			0.080		0.170					
9					0.010					
Length	1.750	1.500	2.000	1.750	2.250	1.750	1.750	1.750	1.750	1.500

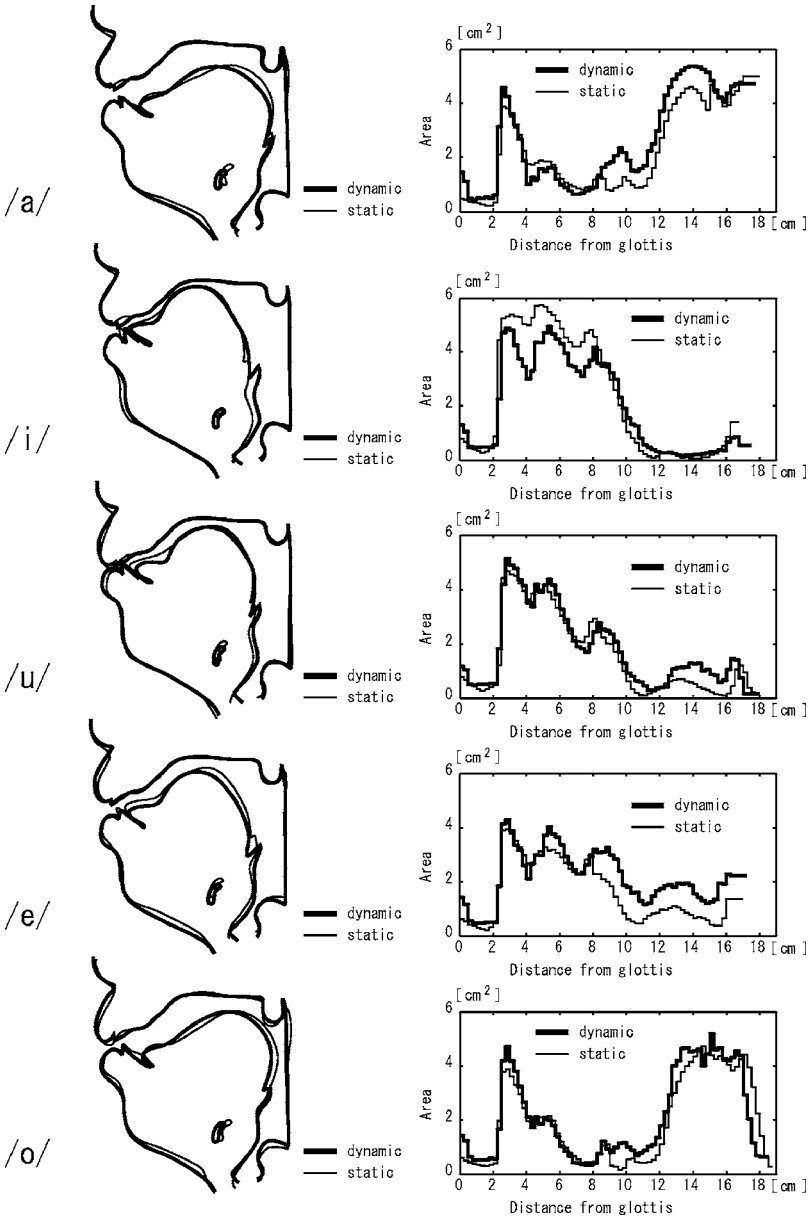


FIG. 12. Left figures indicate the position of speech organs in the midsagittal image, and right figures show area functions from the 3D cine-MRI and static MRI. "Dynamic" represents the 3D cine-MRI data, and "static" denotes the static MRI data, in all the figures.

ACKNOWLEDGMENT

This research was conducted as part of "Research on Human Communication" with funding from the National Institute of Information and Communications Technology.

Adachi, S., and Yamada, M. (**1999**). "An acoustical study of sound production in biphonic singing, Xöömij," J. Acoust. Soc. Am. **105**, 2920–2932.

Alwan, A., Narayanan, S., and Haker, K. (**1997**). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics," J. Acoust. Soc. Am. **101**, 1078–1089.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am. **90**, 799–828.

Caussé, R., Kergomard, J., and Lurton, X. (**1984**). "Input impedance of brass musical instruments—Comparison between experiments and numerical models," J. Acoust. Soc. Am. **75**, 241–254.

Chiba, T., and Kajiyama, M. (**1942**). *The Vowels—Its Nature and Structure* (Tokyo-Kaiseikan, Tokyo).

Dang, J., and Honda, K. (**1997**). "Acoustic characteristics of the piriform fossa in models and humans," J. Acoust. Soc. Am. **101**, 456–465.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague) (2nd ed., **1970**).

Foldvik, A. K., Kristiansen, U., and Kværness, J. (**1993**). "A time-evolving three-dimensional vocal tract model by means of magnetic resonance imaging (MRI)," Proc. Eurospeech **93**, 557–559.

Foldvik, A. K., Husby, O., Kværness, J., Nordli, I. C., and Rinck, P. A. (**1990**). "MRI (Magnetic resonance Imaging) film of articulatory movements," Proc. ICSLP **1**, 421–422.

Foldvik, A. K., Kristiansen, U., Kværness, J., Torp, A., and Torp, H. (**1995**). "Three-dimensional ultrasound and magnetic resonance imaging: A new dimension in phonetic resarch," Proc. 12th ICPhS **4**, 46–49.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–707.

Hiraishi, K., Narabayashi, I., Fujita, O., Yamamoto, K., Sagami, A., Hisada, Y., Saika, Y., Adachi, I., and Hasegawa, H. (**1995**). "Blueberry juice: preliminary evaluation as an oral contrast agent in gastrointestinal MR imaging," Radiology **194**, 119–123.

Honda, K., Takemoto, H., Kitamura, T., Fujita, S., and Takano, S. (**2004**). "Exploring human speech production mechanisms by MRI," IEICE Trans. Inf. Syst. **87**, 1050–1058.

Johansson, C., Sundberg, J., Wilbrand, H., and Ytterbergh, C. (**1983**). "From sagittal distance to area: A study of transverse, cross-sectional area in the pharynx by means of computer tomography," R. Inst. Technol. STL-QPSR **4/1983**, 39–49.

Kitamura, T., Honda, K., and Takemoto, H. (**2005a**). "Individual variation of the hypopharyngeal cavities and its acoustic effects," Acoust. Sci. & Tech. **26**, 16–26.

Kitamura, T., Takemoto, H., Honda, K., Shimada, Y., Fujimoto, I., Shakudo, Y., Masaki, S., Kuroda, K., Oku-uchi, N., and Senda, M. (**2005b**). "Difference in vocal tract shape between upright and supine postures: Observations by an open-type MR scanner," Acoust. Sci. & Tech. **26**, 465–468.

Maeda, S. (**1990**). "Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 131–150.

Masaki, S., Tiede, M. K., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., and Ninomiya, N. (**1999**). "MRI-based speech production study using a synchronized sampling method," J. Acoust. Soc. Jpn. (E) **20**, 375–379.

Mohammad, M., Moore, E., Carter, J. N., Shadle, C. H., and Gunn, S. J. (**1997**). "Using MRI to image the moving vocal tract during speech," Proc. Eurospeech **97**, 2027–2030.

Narayanan, S., Byrd, D., and Kaun, A. (**1999**). "Geometry, kinematics, and acoustics of Tamil liquid consonants," J. Acoust. Soc. Am. **106**, 1993–2007.

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). "An articulatory study of fricative consonants using magnetic resonance imaging," J. Acoust. Soc. Am. **98**, 1325–1347.

Narayanan, S. S., Alwan, A. A., and Haker, K. (**1997**). "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," J. Acoust. Soc. Am. **101**, 1064–1077.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., and Byrd, D. (**2004**). "An approach to real-time magnetic resonance imaging for speech production," J. Acoust. Soc. Am. **115**, 1771–1776.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. (**1999**). "SENSE: sensitivity encoding for fast MRI," Magn. Reson. Med. **42**, 952–962.

Rokkaku, M., Hashimoto, K., Imaizumi, S., Niimi, S., and Kiritani, S. (**1986**). "Measurements of the Three-Dimensional Shape of the Vocal Tract Based on the Magnetic Resonance Imaging Technique," Ann. Bull. RILP. **20**, 47–54.

Sechtem, U., Pflugfelder, P., and Higgins, C. B. (**1987**). "Quantification of cardiac function by conventional and cine magnetic resonance imaging," Cardiovasc. Intervent Radiol. **10**, 365–373.

Shirai, K., and Honda, M. (**1976**). "An articulatory model and the estimation of articulatory parameters by nonlinear regression method," Electron. Commun. Jpn. **55**, 35–43.

Stone, M. (**1990**). "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data," J. Acoust. Soc. Am. **87**, 2207–2217.

Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W. S., and Lundberg, A. (**2001a**). "Modeling the motion of the internal tongue from tagged cine-MRI images," J. Acoust. Soc. Am. **109**, 2974–2982.

Stone, M., Davis, E. P., Douglas, A. S., NessAiver, M., Gullapalli, R., Levine, W. S., and Lundberg, A. J. (**2001b**). "Modeling tongue surface contours from cine-MRI images," J. Speech Lang. Hear. Res. **44**, 1026–1040.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**1998**). "Vocal tract area functions for an adult female speaker based on volmetric imaging," J. Acoust. Soc. Am. **104**, 471–487.

Story, B. H., Titze, I. R., and Hoffman, E. A. (**2001**). "The relationship of vocal tract shape to three voice qualities," J. Acoust. Soc. Am. **109**, 1651–1667.

Takemoto, H., Kitamura, T., Nishimoto, H., and Honda, K. (**2004**). "A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions," Acoust. Sci. & Tech. **25**, 468–474.

Yang, C.-S., and Kasuya, H. (**1994**). "Accurate measurement of vocal tract shapes from magnetic resonance images of child, female and male subjects," Proc. ICSLP **94**, 623–626.

A constitutive model of the human vocal fold cover for fundamental frequency regulation

Kai Zhang and Thomas Siegmund[a)]
School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, Indiana 47907

Roger W. Chan
Otolaryngology – Head and Neck Surgery, and Biomedical Engineering, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9035

(Received 4 April 2005; revised 10 October 2005; accepted 29 November 2005)

The elastic as well as time-dependent mechanical response of the vocal fold cover (epithelium and superficial layer of the lamina propria) under tension is one key variable in regulating the fundamental frequency. This study examines the hyperelastic and time-dependent tensile deformation behavior of a group of human vocal fold cover specimens (six male and five female). The primary goal is to formulate a constitutive model that could describe empirical trends in speaking fundamental frequency with reasonable confidence. The constitutive model for the tissue mechanical behavior consists of a hyperelastic equilibrium network in parallel with an inelastic, time-dependent network and is combined with the ideal string model for phonation. Results showed that hyperelastic and time-dependent parameters of the constitutive model can be related to observed age-related and gender-related differences in speaking fundamental frequency. The implications of these findings on fundamental frequency regulation are described. Limitations of the current constitutive model are discussed. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2159433]

PACS number(s): 43.70.Aj, 43.70.Bk [AL] Pages: 1050–1062

I. INTRODUCTION

In an understanding of voice production processes it is of great interest to obtain predictions of the fundamental frequency of phonation F_0 and to compare such model predictions to experimentally determined trends. Intrinsically, the fundamental frequency of phonation is dependent on the mechanical properties of the tissue in the vocal fold. Specifically, the human vocal fold cover, or superficial layer of the lamina propria, is predominantly an extracellular matrix (ECM) that is optimally designed for tissue vibration and sound production in response to a unique set of biomechanical stimuli in the human larynx. These stimuli include small-strain deformation due to small-amplitude vibration of the vocal fold as well as large-strain deformation due to vocal fold posturing.[1] Small strain deformation occurs at high magnitudes of acceleration (up to 200–300 g) and at high frequencies (around 100–150 Hz for male voice, 200–300 Hz for female voice).[2] Large strain deformation occurs at low frequencies (around 1–10 Hz) and is relevant to vocal fold posturing, i.e., the processes of vocal fold length changes and abduction/adduction creating an optimal set of vocal fold shape and glottal geometry, in order to achieve a desired mode of vibration and vocal fundamental frequency.[1] Vocal fold length adjustments during posturing are primarily driven by the antagonistic actions between cricothyroid and thyroarytenoid muscle activities, resulting in a large strain deformation at tensile strains ranging from 0 to over 40%, or perhaps over 50%. Such high magnitudes of strain are found in situations where singers need to go up the musical scale by more than two octaves. Specific levels of deformation depend on the actual desired fundamental frequency, the relationship between the amplitude of vocal fold vibration and the vocal fold length, and the subglottal pressure.[3] These length changes in the vocal fold are related to the stress level in the vocal fold through the mechanical response of the vocal fold tissue. Furthermore, it is of interest to obtain an understanding of the time course of fundamental frequency changes during speech $F_0(t)$ where t is time. Experimental data on the large-strain mechanical response of vocal fold tissue clearly demonstrate a nonlinear time-dependent behavior.[4] Due to this inherent viscous nature of the vocal fold tissue, the time-dependent changes in vocal fold length are directly linked to changes in fundamental frequency. For an understanding of these phenomena from a continuum mechanics perspective, a constitutive model that reliably describes the stress-strain response of tissues in the vocal fold is of importance. Specifically, the focus of the present study is to investigate fundamental frequency predictions under consideration of a large-strain time-dependent tensile stress-strain response of the vocal fold cover. In the past vocal fold tissue has commonly been described by either linear elastic models,[5,6] nonlinear elastic models,[7–10] or viscoelastic models through the introduction of multiple relaxation times or a time dependence of the elastic constants.[1,11,12] These models possess limitations in dealing with the highly nonlinear tissue response, especially in the large-strain time-dependent response. In the present paper the constitutive description is based on the idea of network models[13] formulated within an

[a)]Author to whom correspondence should be addressed. Electronic mail: siegmund@ecn.purdue.edu

inelastic and time-dependent continuum constitutive relation such that the tissue response to any load is the outcome of an increment solution. For the description of the hyperelastic response of the tissue the Ogden model[14] is used, which has been shown to be promising in the characterization of the tensile behavior of other soft tissues, e.g., brain tissue.[15,16] In addition, a flow rule[17,18] is defined which correlates stress and strain to the magnitude of the time-dependent inelastic deformation. Models of this type have been shown to be capable of describing a wide range of experimentally observed effects in the mechanical behavior of polymers[17–20] and soft tissues.[18,21,22]

The model is applied to results of measurements of the tensile mechanical properties of vocal fold cover specimens obtained from human subjects. Through the constitutive model one can then establish a link between the tissue constitutive properties and histological structure of the vocal fold lamina propria. This provides insights into the relationship between tissue structure and tissue function. Specifically, the present study attempts to correlate the relative densities of collagen and elastin fibers in the vocal fold reported in the literature, including their age-related and gender-related differences,[23,24] with age- and gender-related changes in the parameters of the constitutive model, and finally empirical age- and gender-related dependencies of the fundamental frequency. Even with the number of available human subjects being limited, preliminary conclusions are drawn which correlated well with empirical data on speaking fundamental frequency changes.

II. MECHANICS OF F_0 REGULATION

A. The string model

The ideal string model of phonation[3] provides the starting point for the investigation of the fundamental frequency of phonation. In this model an ideal undamped situation is assumed, and F_0 can be expressed as a function of the current vocal fold length l, the tissue density ρ, and the tissue longitudinal stress σ:

$$F_0 = \frac{1}{2l}\sqrt{\frac{\sigma}{\rho}}. \tag{1}$$

Recently, Titze and Hunter[25] showed that the string model described by Eq. (1) may underestimate the fundamental frequencies of the vocal fold, specifically the vocal ligament, especially at low ranges of strain. By taking into account the effects of variable cross-sectional area and bending stiffness of the vocal fold macula flavae they derived an empirical correction rule to correct the ideal string model as

$$F_0 = \frac{1}{2l}\sqrt{\frac{\sigma}{\rho}}[1 - 0.45\ln(\lambda_u - 1)]. \tag{2}$$

With this correction, the fundamental frequency is dependent on the magnitude of deformation applied due to posturing. Throughout the present study, deformation is measured as stretch λ. In a uniaxial loading condition, stretch is defined as $\lambda_u = 1 + \epsilon = l/L$, with $\epsilon = (l/L) - 1$ being strain and L being the resting length of the vocal fold. Equation (2) was defined to be valid in the range of $\lambda_u = 1.1$ to 1.5.[25]

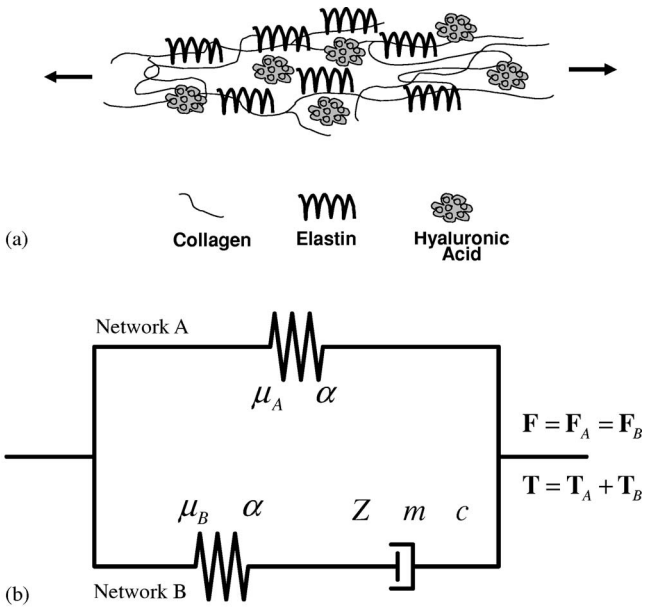


FIG. 1. (a) A schematic representation of molecular networks in the vocal fold extracellular matrix (ECM) following Ref. 26. (b) A one-dimensional rheological representation of the constitutive model, with a hyperelastic equilibrium network (A) in parallel with an inelastic network (B).

In the following, we focus on implications of the constitutive model describing tissue deformation on fundamental frequency regulation. Such effects can be revealed by combining the string models of Eqs. (1) and (2) with an appropriate constitutive model of the vocal fold tissue.

B. A constitutive model for the vocal fold cover

The use of a network-type constitutive model is motivated by the underlying tissue structure. Following Gray *et al.*,[26,27] Fig. 1(a) depicts a schematic representation of the molecular networks present in the ECM of the vocal fold lamina propria. The main constituents include networks of fibrous proteins (collagen and elastin) and water-absorbing, space-occupying hyaluronic acid molecules. Mechanically, the fibrous proteins collagen and elastin provide a structural framework, whereas hyaluronic acid and other interstitial protein molecules contribute to regulate the viscosity and elasticity of the ECM.[26,27] Reflecting this molecular structure, the constitutive model consists of a time-independent equilibrium network (A) in parallel with a time-dependent network (B) [Fig. 1(b)]. No specific assumption is made with regard to collagen, elastin, or interstitial protein molecules exclusively contributing to either network considered in the model. Instead, both the equilibrium response and the inelastic response are assumed to arise from molecular interactions between connected network components, among collagen, elastin, or other protein molecules. In the following, the constitutive model is described in a complete three-dimensional formulation. However, in view of the string model [Eq. (1)], appropriate one-dimensional relationships are given.

The deformation gradient $\mathbf{F}$ relates the length and orientation of a material element $d\mathbf{X}$ in the reference configuration

to its length and orientation in the deformed configuration $d\mathbf{x}$ through $d\mathbf{x}=\mathbf{F}\cdot d\mathbf{X}$. Initial length and current length are defined as $(dL)^2=d\mathbf{X}^T\cdot d\mathbf{X}$ and $(dl)^2=d\mathbf{x}^T\cdot d\mathbf{x}$, respectively, so that stretch λ is

$$\lambda^2=\frac{d\mathbf{x}^T\cdot d\mathbf{x}}{d\mathbf{X}^T\cdot d\mathbf{X}}=\mathbf{N}^T\cdot\mathbf{F}^T\cdot\mathbf{F}\cdot\mathbf{N} \tag{3}$$

in which $\mathbf{N}$ is the unit vector along the orientation of $d\mathbf{X}$.

The two networks (A) and (B) are in parallel with each other such that the applied deformation gradient $\mathbf{F}$ is identical to the deformation gradients in the networks, $\mathbf{F}_A$, $\mathbf{F}_B$:

$$\mathbf{F}=\mathbf{F}_A=\mathbf{F}_B. \tag{4}$$

A one-dimensional rheological model representation of the model is given in Fig. 1(b). Following this parallel arrangement of networks, the total Cauchy stress $\mathbf{T}$ is the sum of the Cauchy stresses in networks (A) and (B), $\mathbf{T}_A$, $\mathbf{T}_B$:

$$\mathbf{T}=\mathbf{T}_A+\mathbf{T}_B. \tag{5}$$

The equilibrium network (A) is characterized by a hyperelastic (large-strain) constitutive behavior. In the analysis of the experimental data it was found that a first-order Ogden model[14] together with the assumption of incompressible material response (with the Poisson's ratio $\nu=0.5$) allows for an appropriate description of the tissue behavior. This model is described by a strain energy density function w of the form[14]

$$w=\frac{2\mu_A}{\alpha^2}(\lambda_1^{\alpha}+\lambda_2^{\alpha}+\lambda_3^{\alpha}-3), \tag{6}$$

where μ_A is the initial shear modulus of the equilibrium network and α is the corresponding dimensionless constant describing the nonlinearity of the elastic response. For incompressible behavior the shear modulus is related to the elastic modulus (Young's modulus, E) with $E=3\mu$. The deformation is characterized through the principal stretches λ_1, λ_2, λ_3 which satisfy $\lambda_1\lambda_2\lambda_3=1$. It is recognized that the vocal fold cover may not be completely isotropic in microstructure and thus in its mechanical properties.[5,11] In the present study, however, effects of anisotropy are not considered.

In the time-dependent network (B), the deformation gradient is decomposed multiplicatively into an elastic component $\mathbf{F}^e$ and a viscoplastic component $\mathbf{F}^p$:

$$\mathbf{F}=\mathbf{F}_B=\mathbf{F}_B^e\cdot\mathbf{F}_B^p. \tag{7}$$

The elastic response of network (B) is again described by an Ogden model, with initial shear modulus μ_B and the power α. Thus, a formulation identical to that of Eq. (6) introduced for the description of network (A) applies for the elastic part of network (B). In the present model it is assumed that the two networks (A) and (B) possess the same power α in the hyperelastic response but different values of the initial network shear moduli, $\mu_A\neq\mu_B$. Following the formulation in Ref. 17 the inelastic rate of the shape change of network (B) is

$$\tilde{\mathbf{D}}_B^p=\dot{\bar{\lambda}}_B^p\frac{\mathbf{S}_B}{\bar{\sigma}_B}=\mathbf{F}_B^e\dot{\mathbf{F}}_B^p(\mathbf{F}_B^p)^{-1}(\mathbf{F}_B^e)^{-1}, \tag{8}$$

where $\mathbf{S}_B$ is the deviator of the Cauchy stress tensor in network (B), and $\bar{\sigma}_B=\sqrt{(3/2)\mathbf{S}_B:\mathbf{S}_B}$ is the corresponding effective stress. The inelastic component of the deformation gradient in network (B) is obtained by inverting the relationship $\tilde{\mathbf{D}}_B^p=\mathbf{F}_B^e\dot{\mathbf{F}}_B^p(\mathbf{F}_B^p)^{-1}(\mathbf{F}_B^e)^{-1}$. The effective inelastic stretch rate $\dot{\bar{\lambda}}_B^p$ depends on the effective inelastic stretch $\bar{\lambda}_B^p$ and the effective stress $\bar{\sigma}_B$:

$$\dot{\bar{\lambda}}_B^p=Z(\bar{\lambda}_B^p-1+\delta)^c(\bar{\sigma}_B)^m. \tag{9}$$

The model introduces three parameters for the description of the inelastic response. The stress exponent m characterizes the dependence of the inelastic deformation on the stress level in network (B). The stretch exponent $c(-1<c<0)$ characterizes the dependence of the rate of inelastic deformation on the current magnitude of inelastic deformation. The viscosity scaling constant Z defines the absolute magnitude of the inelastic deformation. Furthermore, δ is a small positive number ($\delta<0.001$) introduced to avoid singularities in the inelastic stretch rate when the inelastic stretch is close to one.[18]

For uniaxial loading of a specimen of initial length L to a current length l the principal stretches are $\lambda_1=\lambda_u$, $\lambda_2=\lambda_3=1/\sqrt{\lambda_u}$ where $\lambda_u=1+\epsilon_u=l/L$. For the equilibrium response as given by the nominal stress in network (A), the total nominal stress $\sigma=\sigma_A$. This stress is obtained as the derivative of the strain energy density w with respect to the applied stretch λ_u:

$$\sigma=\sigma_A=\frac{\partial w}{\partial\lambda_u}=\frac{2\mu_A}{\alpha}(\lambda_u^{\alpha-1}-\lambda_u^{-(1/2)\alpha-1}). \tag{10}$$

Under consideration of the time-dependent response, stress is calculated from the elastic components of stretch λ_A^e, λ_B^e and both depend on the applied stretch λ_u and the applied stretch rate $\dot{\lambda}_u$. For network (A) the elastic stretch rate is $\dot{\lambda}_A^e=\dot{\lambda}_A=\dot{\lambda}_u$ and the elastic stretch is $\lambda_u=\lambda_A=\lambda_A^e$. In network (B) the elastic stretch rate $\dot{\lambda}_B^e$ can be expressed as

$$\dot{\lambda}_B^e=\dot{\lambda}_u\frac{\lambda_B^e}{\lambda_u}-\frac{2}{3}\lambda_B^e\cdot Z\cdot\mathrm{sgn}(\lambda_B^e-1)\cdot\left\{\sqrt{\frac{1}{3}\left[\left(\frac{\lambda_u}{\lambda_B^e}\right)^2+\frac{2\lambda_B^e}{\lambda_u}\right]}-1+\delta\right\}^c\cdot\left|\frac{2\mu_B}{\alpha}[(\lambda_B^e)^{\alpha}-(\lambda_B^e)^{-\alpha/2}]\right|^m \tag{11}$$

with the elastic stretch component λ_B^e found through numerical integration[28] of Eq. (11). With these formulations, the total nominal stress σ can be obtained:

$$\sigma=\frac{2\mu_A}{\alpha}[\lambda_u^{\alpha-1}-\lambda_u^{-\alpha/2-1}]+\frac{2\mu_B}{\alpha}[(\lambda_B^e)^{\alpha-1}-(\lambda_B^e)^{-\alpha/2-1}]. \tag{12}$$

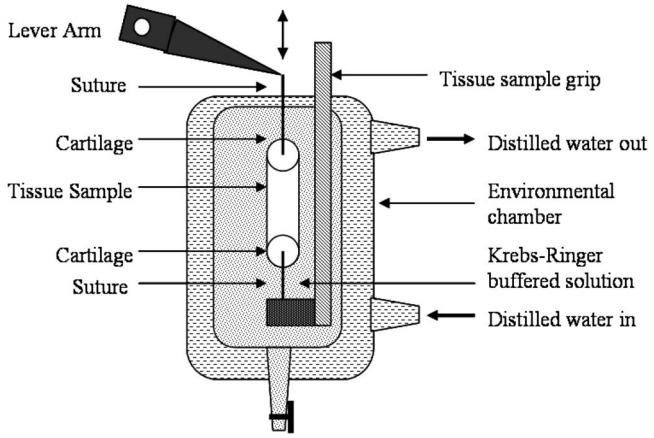


FIG. 2. Schematic of the experimental setup for uniaxial tensile stretch-release test of vocal fold tissues using a dual-mode servo-control lever system.

III. TISSUE DATA

A. Experimental method

The passive uniaxial tensile stress-strain response of the human vocal fold cover was measured by sinusoidal stretch-release (loading-unloading), with the use of a dual-mode servo control lever system (Aurora Scientific Model 300B-LR, Aurora, ON) (Fig. 2). The servo control lever system allowed for precise real-time measurements of the displacement and force of the lever arm, with a displacement accuracy of 1.0 μm and a force resolution of 0.3 mN. The servo-control lever system possessed a displacement range of up to 8–9 mm in the frequency range of 1–10 Hz. The experimental protocol was approved by the Institutional Review Board of UT Southwestern Medical Center.

The sinusoidal tensile stretch-release protocol for obtaining stress-strain data was similar to those of Alipour and Titze[7] and Perlman,[29] but the dissection was different because only canine vocal fold tissues were tested in those previous studies, instead of human. In the present study, the vocal fold cover, i.e., the epithelium and the superficial layer of the lamina propria, was dissected with instruments for phonomicrosurgery, as summarized in Chan and Titze[11] (please note that in Chan and Titze[11] the vocal fold cover was referred to as the "mucosa"). In particular, the superficial layer of the lamina propria was separated from the intermediate layer (the vocal ligament) and the thyroarytenoid muscle (the vocal fold body) by a Bouchayer spatula for blunt dissection. This lateral boundary of excision was facilitated by the natural plane of dissection between the superficial layer and the middle layer of the lamina propria, which was correlated to elastin fiber density.[30] No muscle tissue was observed in all of the vocal fold cover specimens isolated. Each cover specimen remained naturally attached to a small portion of thyroid cartilage at the anterior end and a small portion of arytenoid cartilage at the posterior end, to which sutures were attached for mounting and for elongation. It is believed that the tissue primarily responsible for bearing stress and undergoing deformation was the cover itself (epithelium and/or superficial layer of the lamina propria), since the vocal ligament and the vocal fold body (muscle) were excluded from the specimen after dissection.

Each vocal fold cover specimen to be tested was mounted vertically to the lever arm through sutures in Krebs-Ringer solution at a *p*H of 7.4 and at 37 °C in a glass environmental chamber. In all experiments the loading rate was 1 Hz. This frequency for vocal fold length changes was selected based on the following considerations. A broad range of speeds of vocal pitch changes was observed in human subjects[31,32] and in canines *in vivo*.[33] A response time of around 50–100 ms was found for human singers and untrained subjects as they attempted ascending and descending pitch jumps at maximum speeds.[31,32] A time constant of about 30 ms was found for exponential increase or decrease in vocal fold strain in dogs under supramaximal electrical stimulation of the superior laryngeal nerve or the recurrent laryngeal nerve, up to a maximum positive strain (elongation) of about 45%.[33] The exponential time constant observed in the canine study was consistent with the human response time, as it was defined as reaching 66.7% of the target pitch, while it would take about three time constants (90 ms) to approach steady state.[33] A time constant of 50 ms would correspond to a loading rate of 10 Hz for 50% vocal fold elongation under sinusoidal deformation, whereas 500 ms would correspond to a loading rate of 1 Hz. Since the vocal fold length changes in these experiments were executed at maximum speeds or supramaximally, for lower speeds of pitch changes such as those during speech, particularly for males, the time constant could approach 500 ms, i.e., a loading rate of 1 Hz as chosen in the present experiments.

Displacement and tensile force of the specimen under cyclic stretch-release were measured, and data were acquired at a sampling rate of 1000 samples per second per channel with a 14-bit signal amplitude resolution. From the vocal fold *in situ* length (resting cadaveric length) L and the current length l the uniaxial stretch of a specimen was obtained as $\lambda_u = l/L$. The nominal tensile stress σ was determined from the measured force divided by the specimen cross-sectional area A, estimated by $A = M/(\rho L)$ where M is the mass of the specimen.[7] The mass of each cover specimen was measured after the completion of the tensile stretch-release experiment. A possible hydration-induced change in mass of the specimens in the present study was not examined, but results of an animal model showed that the mass of the porcine vocal fold mucosa increased by around 5%–10% with an hour of hydration in an isotonic solution (incubation in Krebs-Ringer solution at *p*H 7.4). This discrepancy in specimen mass could indeed introduce a larger variability into the stress-strain data, as mass affects the calculation of cross-sectional area of the specimen. However, based on our animal data the increase in variability was limited to only about 5% of the specimen mass before incubation, resulting in a 5% or less error in the calculated stress values. The nominal stress was selected as a stress measure in order to facilitate the comparison to previously published data on the vocal ligament and the vocal fold body (muscle).[8] This definition of stress is also commonly used in other studies in soft tissue biomechanics.[15,16,21] Each specimen was loaded to a

constant value of maximum length at load reversa l_{rev} such that the stretch at load reversal is defined as $\lambda_{u,rev}=l_{rev}/L$.

Vocal fold cover specimens were dissected from 11 larynges excised within about 24 h postmortem, procured from autopsy from human cadavers free of head and neck disease and laryngeal pathologies. All subjects were nonsmokers and were Caucasians, although race was never a factor in the procurement. Tissue samples were obtained from six male larynges of ages $Y=17$, 33, 51, 65, 66, and 99 years. The vocal fold *in situ* lengths of the male specimens were $L=14.1$, 17.8, 22.4, 18.0, 20.4, and 17.0 mm, respectively, in the order of increasing subject age. The *in situ* lengths of male specimens did not significantly depend on age ($p=0.58$). Exploiting the displacement range of the lever system the male specimens were loaded to stretches of up to $\lambda_{u,rev}=1.35$. Samples were obtained also from five female larynges of ages $Y=80$, 82, 83, 85, and 97 years. The vocal fold *in situ* lengths of the female specimens were $L=14.5$, 14.9, 15.1, 13.7, and 13.9 mm, respectively, in the order of increasing subject age. The *in situ* lengths of female specimens did not significantly depend on age ($p=0.31$). Female specimens possess shorter *in situ* lengths than male specimens at a level of statistical significance ($p=0.02$). Exploiting the displacement range of the lever system, the female specimens were loaded to stretches of up to $\lambda_{u,rev}=1.5$ except for the specimen $Y=97$ which was tested to smaller maximum deformation.

In the experiments, the goal was to prevent any slackness in the sutures and to mount the specimen at a length as close to its *in situ* length as possible. No further predeformation was used in the experiments. The *in situ* length was used as the reference length for calculation of stretch values. Mounting was accomplished such that a tensile force of 0.01–0.02 N was applied to the specimen. This force relaxed until the start of each test. The stress-stretch curves reported are those after preconditioning, i.e., after the stretch-release loading had reached a stable response. A stabilized response was typically achieved after 10 to 12 cycles. During preconditioning the experimental data shifted from the initial zero position of the load-elongation diagrams due to creep deformation. For the analysis of the experimental data in the preconditioned state and their characterization by the constitutive model the creep offset deformation was subtracted and load-elongation curves were placed at zero.

Measured tensile stress-stretch curves obtained for the male and female specimens are given in Fig. 3. While only a limited number of specimens are available, all curves show similar characteristics. A significant hysteresis and a strong nonlinear dependence of stress on the applied stretch is observed, similar to previous stress-strain data on the human vocal ligament.[8] For higher stretch values the fit is still good but less perfect. This is attributed to potential heterogeneous deformation at larger magnitudes of deformation.

B. Constitutive characterization

The constitutive model is applied to characterize each of the specimens with the multi-parameter constitutive model. The resulting constitutive material parameters $\mu_A, \alpha, \mu_B, Z, m, c$ obtained for the individual specimens are subsequently investigated in their age and gender differences. The Appendix describes the procedure to determine the values of the constitutive parameters.

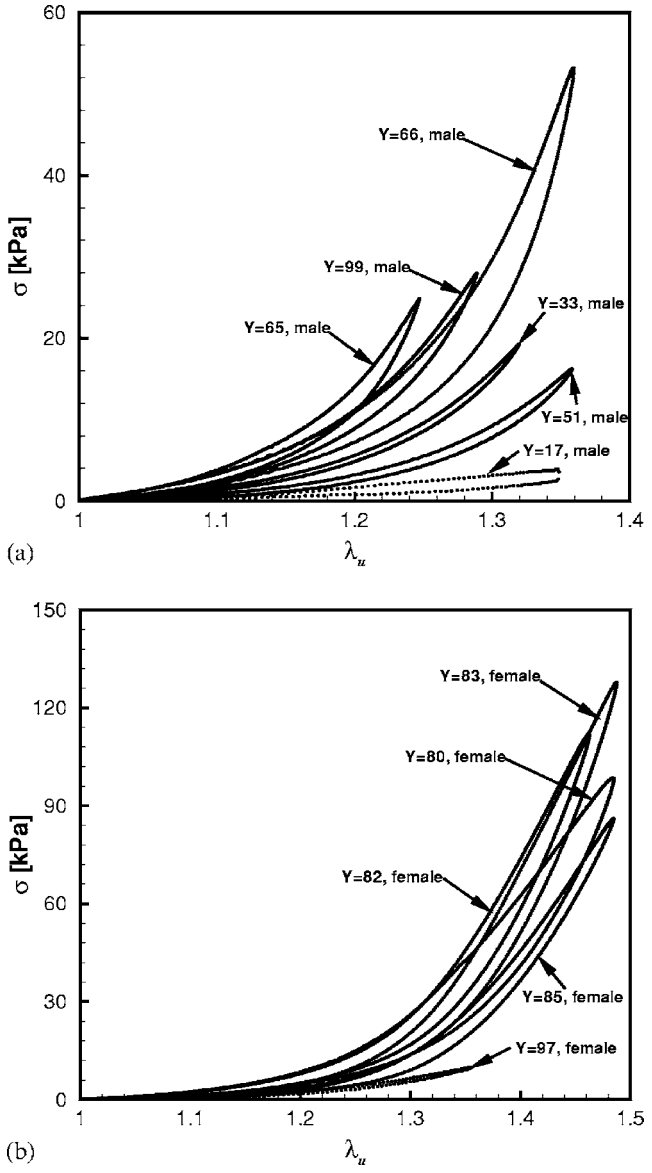


FIG. 3. Experimental tensile stress-stretch response at loading frequency of 1 Hz: (a) male vocal fold cover specimens and (b) female vocal fold cover specimens (Y=age in years).

Figure 4 shows examples of specimen response in the constitutive model in comparison to experimental data for two male and two female samples, respectively. The two male examples $Y=17$ and 99 are chosen to document in detail the differences in tissue response between a very young subject and one of advanced age, and to demonstrate the flexibility of the constitutive model to characterize significantly different types of stress-stretch response in a unified framework. It can be seen from Fig. 4(a) that the samples for the male subjects $Y=17$ and 99 exhibit quite significantly different characteristics with respect to hysteresis and the stress level associated with an applied level of deformation. Such differences are interpreted as possibly age related, due to the ultrastructural changes occurring with age,[23,24] as in

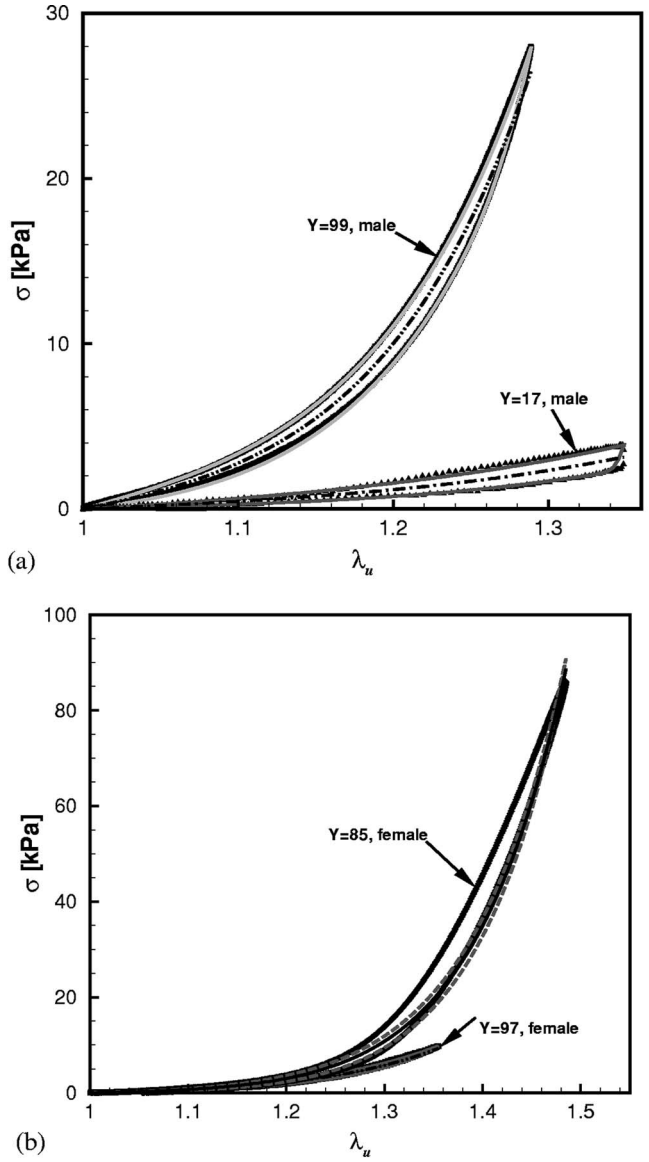


FIG. 4. Comparison between experimental data (———) and simulation results (———) of tensile stress-stretch response at 1 Hz: (a) male specimens $Y=17$ and $Y=99$ and (b) female specimens $Y=85$ and $Y=97$.

some other human soft tissues such as lung tissue.[34] Despite these differences the constitutive model can characterize both stress-stretch curves well. Parameter values of the constitutive parameters for all male subjects are summarized in Table I. For all samples the experimental data and the model characterization were highly correlated ($R^2 \geqslant 0.9792$), indicating the capability of the constitutive model to describe the stabilized hysteretic tissue response, at least among the limited number of specimens examined. The two female samples depicted in Fig. 4(b), $Y=97$ and 85, are chosen due to the difference in the maximum applied stretch level, $\lambda_{u,rev}=1.35$ for $Y=97$ and $\lambda_{u,rev}=1.48$ for $Y=85$. For the 97-year-old specimen, a good agreement between experimental data and constitutive description is obtained over the entire range of stretch, with $R^2=0.9976$. For the female specimen $Y=85$, and similarly for all other female specimens, the hysteretic response at large stretches, $\lambda_u > 1.3$, is not as well approximated as for moderate values of stretch, $\lambda_u < 1.3$. While the peak stress at load reversal is well predicted, the current model tends to underpredict the magnitude of the hysteresis observed experimentally. The parameter values of the constitutive model for the case of female subjects are also summarized in Table I. The R^2 values indicate a good approximation of the experimental data by the constitutive model, especially for $\lambda_u < 1.3$.

C. Age- and gender-related differences in tissue behavior

For male specimens it was possible to establish a relationship between the tissue response and age. A nonlinear regression analysis[35] was conducted for the parameter values of the equilibrium network (A) μ_A and α. Both data sets could be fitted with exponential functions: $\mu_A(Y)=7.0[1.0-\exp(-0.018\,84Y)]$, with the regression coefficient $R^2=0.615$, and $\alpha(Y)=15.35[1.0-\exp(-0.058\,41Y)]$, with $R^2=0.8744$. The individual data for μ_A and α, together with the exponential regression functions and the corresponding 95% confidence intervals, are given in Figs. 5(a) and 5(b). The value of μ_A, i.e., the stiffness of the tissue equilibrium response, increases with age, with the increase more significant in the age range below 65 years of age and a saturation at older age. The regression data suggest that μ_A for older males is up to three times larger than that of young adult males. Similarly, the nonlinearity of the tissue hyperelastic response, as characterized by the power α in the Ogden model, also increases with age [Fig. 5(b)]. This effect seems to be more significant in the age range below 50 years of age and again appears to saturate at older age. The present data suggest that the value of α increases by approximately a factor of 2 from young adult males to old males. The age

TABLE I. Summary of the parameters of the constitutive model and R^2 values for comparison of experimental data and constitutive description.

	Male specimens						Female specimens				
Age	17	33	51	65	66	99	80	82	83	85	97
μ_A (kPa)	1.37	3.50	1.765	6.67	5.6	6.39	4.117	3.63	1.307	1.426	1.77
α	8.8	14.1	14.6	16.1	14.5	14.4	14.54	15.44	19.47	16.83	12.5
μ_B (kPa)	80	20	25	60	130	40	190.0	70.0	180.0	60.0	10.0
m	1.5	1.8	1.2	1.6	1.6	2.1	1.1	1.7	1.0	1.0	1.3
c	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−1.0	−0.7
$Z[s^{-1}(\text{kPa})^{-m}]$	6.0e-2	9.0e-2	5.0e-2	6.5e-3	3.4e-3	1.0e-2	8.0e-3	1.0e-2	8.0e-3	3.5e-2	2.0e-1
R^2 ($\lambda_u < 1.3$)							0.991	0.990	0.889	0.985	0.993
R^2 (entire curve)	0.979	0.999	0.999	0.999	0.998	0.999	0.908	0.991	0.520	0.987	0.997

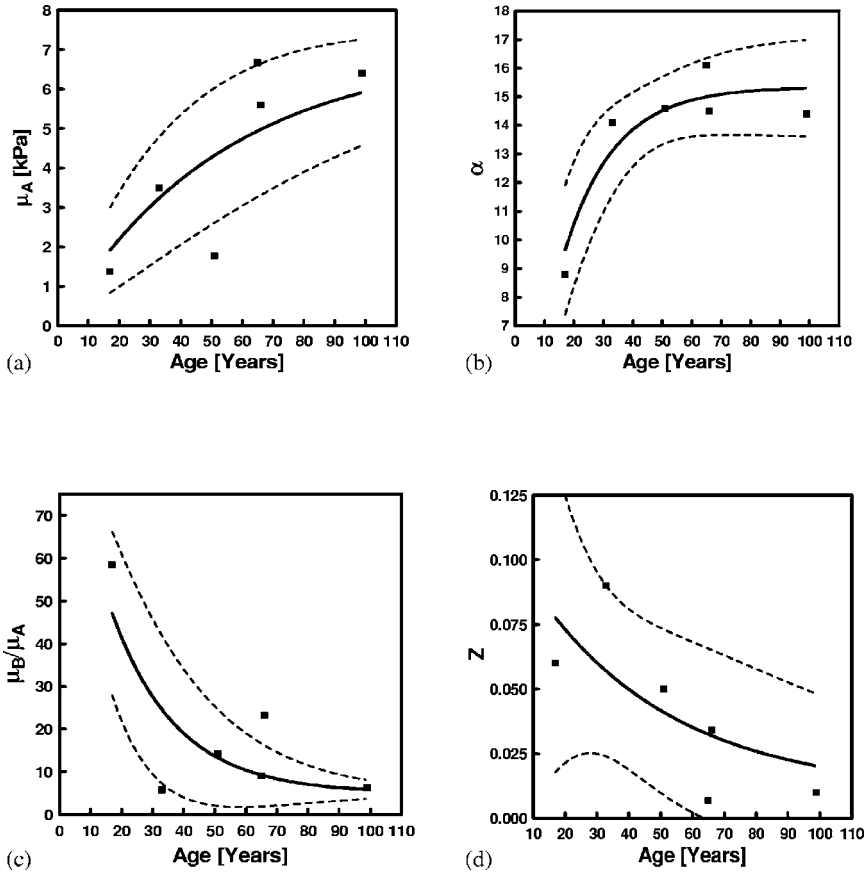


FIG. 5. Age dependence of the model parameters for the male specimens: (a) initial shear modulus of the equilibrium network (μ_A), (b) power in the Ogden model (α), (c) ratio between the stiffness of the time-dependent network relative to the stiffness of the equilibrium network (μ_B/μ_A), and (d) the viscosity scaling constant (Z). Dotted lines are the 95% confidence intervals of the curve-fitting exponential functions.

dependence of the ratio between the initial shear modulus of the time-dependent network (B) μ_B relative to the initial shear modulus of the equilibrium network (A) μ_A is given in Fig. 5(c) together with a nonlinear regression function and the corresponding 95% confidence intervals. The exponential function $\mu_B/\mu_A = 95.0\exp(-0.047\,85Y) + 5.0$ provides a reasonable fit to the experimental data with a regression coefficient of $R^2 = 0.6625$. It should be noted here again that the same value of α was used for both network (A) and network (B) in the model. For the dependence of the effective inelastic stretch rate on the effective inelastic stretch, a single value of the parameter c was used throughout the study, i.e., $c = -1.0$, and hence there was no age dependence for this parameter. The stress dependence of the inelastic deformation as expressed through the power m was found to not significantly vary with age. The mean value of this parameter was determined as $\bar{m} = 1.633$ with a standard deviation of 0.3011. A one-sample two-tailed t test was conducted to verify if the stress dependence of the viscoplastic flow law was significantly different from $m = 0$. This test resulted in a p value of $p < 0.0001$, demonstrating statistical significance. The finding that m can be assumed as a constant is especially interesting when considering the analysis of the age dependence of the viscosity scaling constant Z. A comparison of the values of this parameter becomes meaningful if m can be treated as a constant, such that all Z values possess the same units, i.e., $\sim[\mathrm{s}^{-1}(\mathrm{kPa})^{-1.6}]$ in the present case. Figure 5(d) depicts the age dependence of Z together with the nonlinear regression function and the corresponding 95% confidence intervals. The nonlinear regression analysis led to an exponential function of $Z(Y) = 0.0998\exp(-0.0229 \cdot Y) + 0.01$, with $R^2 = 0.5504$. Figure 5(d) suggests that the viscous response of the tissue, as characterized by Z, decreases significantly with age. For older males Z is found to be approximately one-third of that of young adult males.

For the female specimens no age dependence could be determined due to the small spread in age in the specimens available. Nevertheless, the dependence of the constitutive parameters on gender is investigated. For female subjects the mean age is $\bar{Y} = 85$ compared to $\bar{Y} = 55$ for the male subject group. In order to examine gender-related effects in subjects of a similar age range, the two youngest male specimens were excluded from the analysis, resulting in $\bar{Y} = 70$ for the male subgroup. Following the results of Fig. 5, in this subgroup the age dependence of the model parameters is then much less pronounced than for the overall male population. For the female specimens a mean value of the tissue equilibrium stiffness of $\bar{\mu}_A = 2.45$ kPa with a standard deviation of 1.32 kPa was obtained, whereas for the male specimens ($\bar{Y} = 70$) it was $\bar{\mu}_A = 5.11$ kPa with a standard deviation of 2.27 kPa. These data result in a male-to-female difference in the mean values of tissue equilibrium stiffness of $\bar{\mu}_A(\mathrm{male}) - \bar{\mu}_A(\mathrm{female}) = 2.66$ kPa. A one-tailed t test was conducted to investigate if this difference between male and female tissue stiffness was statistically significant. This test resulted in a value of $p = 0.031$ (DOF = 4.591), indicating that this gender-related difference in stiffness was significant. In addition, a nonparametric Mann-Whitney U test was conducted to further investigate if the finding of $\bar{\mu}_A(\mathrm{male}) > \bar{\mu}_A(\mathrm{female})$ was

statistically significant, without the assumption of normal sample distribution. This test returned a value of $p=0.0557$, close to statistical significance. A similar analysis of gender-related difference was also conducted for the power of the hyperelastic response in the Ogden model. For the female specimens a mean value of $\bar{\alpha}=15.76$ with a standard deviation of 2.60 was obtained. For the male specimens ($\bar{Y}=70$) it was found that $\bar{\alpha}=14.90$ with a standard deviation of 0.4. These data result in a difference between male and female of $\bar{\alpha}(\text{male})-\bar{\alpha}(\text{female})=-0.856$. Results of both the t test and Mann-Whitney U test indicate that the difference was not statistically significant with $p>0.5$ in both cases.

For the parameters characterizing the inelastic tissue response no further statistically significant gender difference was observed: for μ_B/μ_A a one-tailed t test for significance in gender resulted in $p=0.093$ (DOF=4.207); the corresponding U test for $\mu_B/\mu_A(\text{male})>\mu_B/\mu_A(\text{female})$ resulted in $p=0.143$. For m a one-tailed t test for significance in gender resulted in $p=0.063$ (DOF=5.733), and the corresponding U test for $m(\text{male})>m(\text{female})$ resulted in $p=0.087$. For Z a two-tailed t test for significance in gender resulted in $p=0.415$ (DOF=4.67), and the corresponding U test for $Z(\text{male})<Z(\text{female})$ resulted in $p=0.2293$.

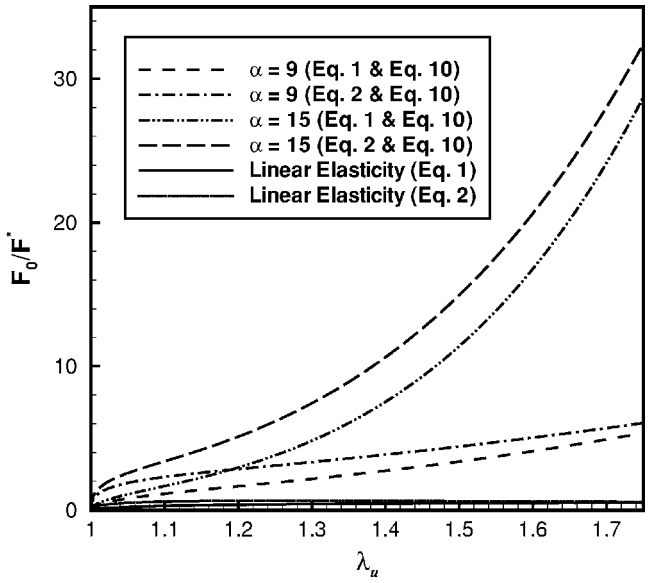


FIG. 6. Dependence of the predicted normalized fundamental frequency F_0/F^* on applied tensile stretch λ_u based on Eqs. (10) and (16). Predictions are shown for a young adult male ($\alpha=9.0$), an old adult male ($\alpha=15.0$), and a linear elastic model (Hooke's law).

IV. ANALYSIS OF THE EQUILIBRIUM F_0

Predictions of fundamental frequency can be obtained from Eqs. (1) and (2) once a constitutive model is established and its parameter values are determined. For comparison purposes, the example of linear elasticity as described by Hooke's law is considered first. Combining a one-dimensional description of linear elasticity model, $\sigma=E\epsilon=E(\lambda_u-1)$, with Eq. (2) one obtains

$$F_{0,\text{lin}}=F^*_{\text{lin}}\frac{\sqrt{\lambda_u-1}}{\lambda_u}[1-0.45\ln(\lambda_u-1)] \tag{13}$$

with

$$F^*_{\text{lin}}=\frac{\sqrt{E}}{2L\sqrt{\rho}}. \tag{14}$$

The fundamental frequency F_0 is thus given as the product of a deformation measure and a reference frequency F^*_{lin} containing the information of the tissue elastic response, its resting length and density.

For the hyperelastic tissue response the combination of Eq. (2) with Eq. (10) leads to a prediction of $F_{0,\text{hyper}}$ in dependence of the constants of the hyperelastic constitutive model and the tensile deformation as

$$F_{0,\text{hyper}}=F^*_{\text{hyper}}\sqrt{\lambda_u^{\alpha-3}-\lambda_u^{-\alpha/2-3}}[1-0.45\ln(\lambda_u-1)] \tag{15}$$

with

$$F^*_{\text{hyper}}=\frac{\sqrt{2\mu_A}}{2L\sqrt{\rho\alpha}}. \tag{16}$$

The influence of vocal fold deformation on fundamental frequency is assessed by plotting the normalized fundamental frequency F_0/F^* as a function of vocal fold length changes as characterized by stretch λ_u (Fig. 6). For the linear elastic model the predicted values of F_0/F^* are only slightly dependent on the applied stretch and become nearly constant for $\lambda_u>1.2$, independently of whether the correction proposed by Titze and Hunter[25] is included or not. This is in significant contrast to results of *in vivo* observations of human vocal fold length during pitch changes.[36] Much better qualitative agreement with the observations of Ref. 36 is obtained with the predictions of F_0/F^* from the hyperelastic model. F_0/F^* depends nonlinearly on λ_u and strongly increases with vocal fold stretch. The change in the value of the parameter α significantly influences the dependence of F_0/F^* on λ_u. For larger values of α ($\alpha=15.0$ typical of subjects of advanced age in the male group) F_0/F^* depends more strongly on stretch than for low values of α ($\alpha=9.0$ typical of young males). Again, these trends are independent of whether the correction function of Titze and Hunter[25] is included or not.

The predicted dependence of F_0 on age for the male specimens is depicted in Fig. 7(a) with F_0 calculated through Eq. (15). For the calculation of F_0, individual specimen resting length data were combined with the constitutive parameters given in Table I. A stretch of $\lambda_u=1.2$—typical length changes in males—and a tissue density of $\rho=1040\ \text{kg/m}^3$ were assumed.[5,36] To put predicted F_0 data into perspective, Fig. 7(a) also shows trend-lines approximating the life-span changes of F_0 based on human speaking F_0 data obtained by Brown *et al.*[37] and Krook[38] A linear regression analysis was conducted, and it was found that the predicted F_0 could be fitted by $F_0(Y)=0.9599Y+42.454$, with the regression coefficient $R^2=0.55$. The present finding of an age-related increase in fundamental frequency due to changes in the hyperelastic constitutive parameters is consistent with empirical life-span changes after middle age where F_0 gradually increases with age in males above $Y\approx 40$.[37–39]

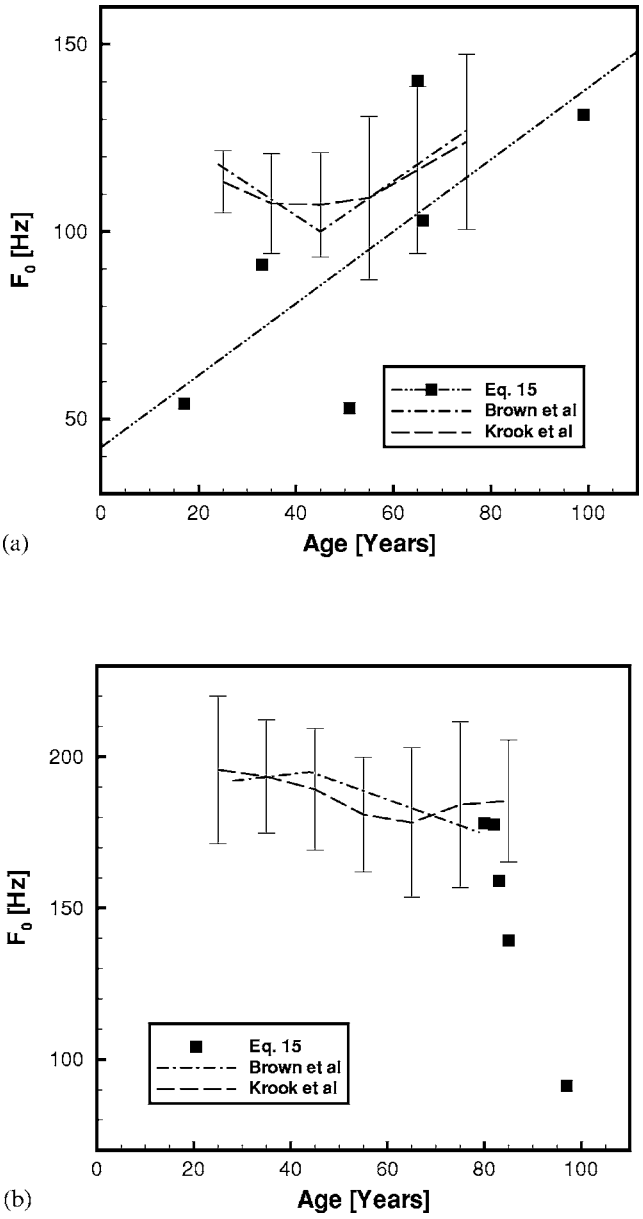


FIG. 7. Dependence of the predicted fundamental frequency F_0 on age for (a) male specimens and (b) female specimens. Data points represent individual specimens of the present study. Two trend-lines depict approximately the change in speaking fundamental frequency with age as reported by Brown *et al.*[37] and Krook[38]

In females, the data of Brown *et al.*[37] and Krook[38] show a gradual decrease of F_0 on age, especially for Brown *et al.*[37] They also found F_0 to be larger in females than in males, as expected. Equation (15) was also used to predict F_0 values for the present female subjects. Again, for this analysis individual specimen resting length data were combined with the constitutive parameters given in Table I, and a tissue density of $\rho=1040\ \mathrm{kg/m^3}$ was assumed. However, for the female subjects a larger value of stretch due to posturing was used, $\lambda=1.3$.[36] Figure 7(b) depicts trends of F_0 from Refs. 37 and 38 in comparison to predicted F_0 values. For several of the subjects, the predicted F_0 is well in the range of data from Refs. 37 and 38, while for the oldest female subject ($Y=97$) the predicted F_0 value is much lower.

In general, the values of F_0 predicted in the present study tend to be lower than those from speaking data. Several factors are deemed responsible for this. First, the inelastic tissue properties need to be considered. For a discussion of this aspect, see the following section. Second, only the vocal fold cover is investigated in the present study. The use of Titze and Hunter's[25] correction factor is only a rough approximation for our description of the vocal fold cover, since it was derived for the vocal ligament. The effect of bending stiffness is expected to be similar, but not the effect of the macula flavae, which are not part of the vocal fold cover. Future studies should include a correction factor more specific to the vocal fold cover, as well as the key contributions of the vocal ligament, which should allow the model to predict more physiological F_0 values. Third, intrinsic laryngeal muscle activities were not included in the present study. Activities of the cricothyroid and the thyroarytenoid muscles could either increase or decrease the stiffness of the vocal fold cover depending on a balance of their antagonistic actions, and it is well documented that they combine to regulate F_0.[39,40]

V. ANALYSIS OF THE NONEQUILIBRIUM F_0

In the results presented in the preceding section the analysis of F_0 was based only on the hyperelastic contribution of the equilibrium network (A). If Eq. (15) was calculated with $\mu_A+\mu_B$ instead of μ_A only, values of F_0 much larger than physiological data would be predicted. For example, for males $Y>50$ an average value of $\bar{F}_0=388$ Hz is obtained at $\lambda_u=1.2$. This finding suggests that the elastic component of network (B) may only partially be contributing to the regulation of F_0. The exact relative amount of elastic and inelastic deformation in network (B) depends on the frequency and the magnitude of the stretch during vocal fold length changes. Predictions of the time-dependent evolution of F_0 can be obtained by combination of Eq. (2) with numerical predictions of the longitudinal tissue stress by use of solutions from Eq. (12). One example of a predicted time dependence of F_0 is given in Fig. 8 where a deformation history of $\lambda_u=1.0\rightarrow 1.2\rightarrow hold\rightarrow 1.1\rightarrow hold\rightarrow 1.2\rightarrow hold\rightarrow 1.1$ is considered. The stretch rate in the computation is $0.4/\mathrm{s}$ such that the initial stretch takes 0.5 s corresponding to a loading frequency of 1 Hz. Such rates of stretch are representative of vocal fold length changes during speech.[31–33] The present model predicts changes in F_0 to occur over short time periods (several ms to several hundred ms) after stretch was changed. If stretch is increased from the previous level, F_0 drops from a higher value to a steady state. If stretch is decreased from the previous level, F_0 increases from a lower value to a steady state. The findings of predicted F_0 changes to occur in the range of up to several hundred milliseconds are in general agreement with exponential time constants for vocal fold length changes if they occurred at submaximal speeds.[33]

In order to quantify the time-dependent response, we define a characteristic time constant as the time for 90% of stress relaxation to have occurred. Figures 9(a) and 9(b) show predicted time constants for male and female speci-

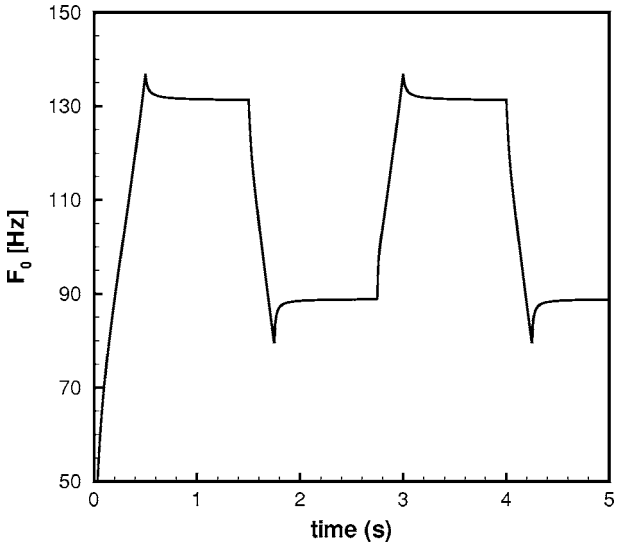


FIG. 8. Predicted time course of F_0 for a stretch history $\lambda_u=0\rightarrow1.2\rightarrow1.1\rightarrow1.2\rightarrow1.1$ and a loading frequency of 1 Hz. Simulation conducted with constitutive parameters of the 99-year-old male specimen with ($Y=99$).

mens, respectively, in dependence of applied stretch. The predicted time constants are found to be dependent on stretch. As stretch increases the inelastic response of the tissue is reduced and the time required for stress relaxation increases. For male subjects the characteristic time constant at $\lambda_u=1.2$ increases with age, $\tau=1.2\times10^4+4.0[1-\exp(-0.00011\times Y)]$, $R^2=0.605$. However, no statistically significant difference between gender was found.

VI. DISCUSSION

The present study documents the large-strain, hyperelastic, and viscous (hysteretic) response of the human vocal fold cover under uniaxial tensile deformation. Experimental stress-stretch (or stress-strain) data were obtained from the superficial layer of the lamina propria (the vocal fold cover), as compared to previous data on the deeper layers of the lamina propria (the vocal ligament).[8] Both the epithelium and the superficial layer of the lamina propria likely contributed to the observed tensile mechanical response, although it was not possible to determine their exact contributions in the present study. In agreement with previous studies on vocal fold elasticity, it was found that the vocal fold cover was highly nonlinear, hysteretic, and with a large interindividual variability, although these observations were based on only a limited number of specimens. To obtain a more fundamental understanding of the structural basis of the tissue deformation response based on age- and gender-related differences, a constitutive modeling approach was employed in the analysis of the data. The constitutive model was formulated based on the application of Ogden's hyperelastic model within an incremental inelastic (viscoplastic) network model. In this constitutive formulation, no arbitrary separation into a low-strain, linear elastic region and a large-strain, nonlinear exponential regime is required, as was common in earlier studies of vocal fold tissue elasticity.[7,8] Results showed that the constitutive model approach was capable of simulating the vocal fold cover deformation response if the applied deformation remained below a stretch of $\lambda<1.3$. For higher values of the applied stretch it appears that more than one inelastic mechanism of deformation needs to be accounted for, particularly for characterizing the female specimens. An extension of the current model to allow for its application at large deformation can be constructed by adding a second inelastic network to the model, for example. Such an approach is currently under investigation.

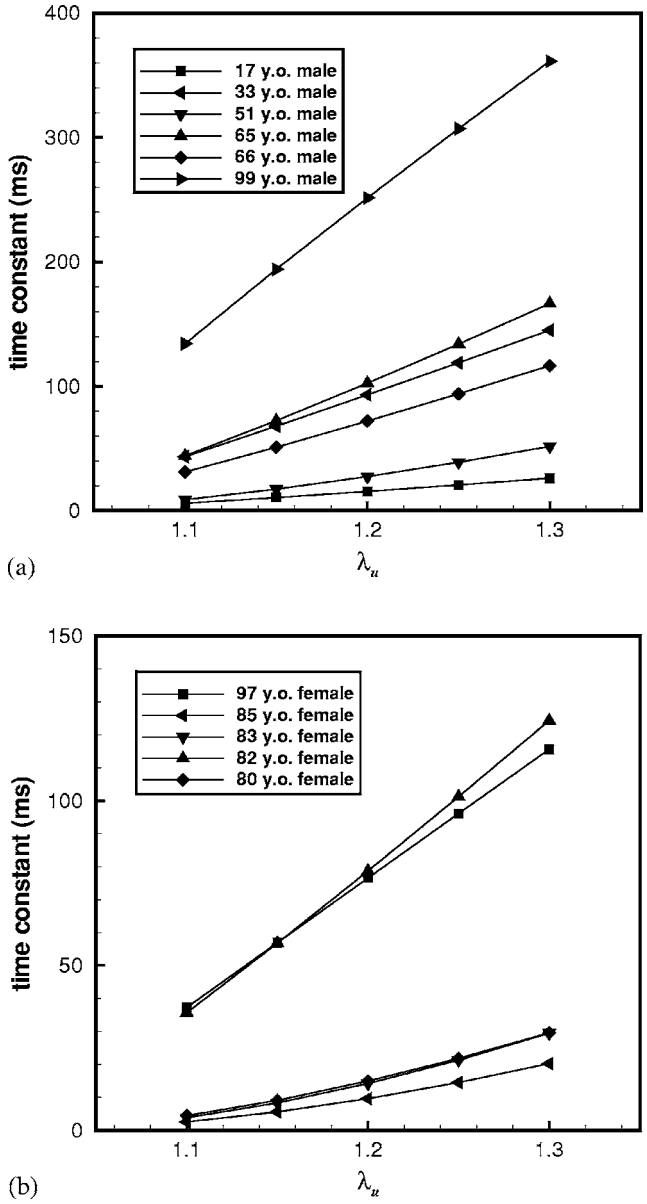


FIG. 9. Predicted dependence of the characteristic time constant of relaxation on the applied stretch for (a) male specimens and (b) female specimens. Loading frequency=1 Hz.

From a functional biomechanical perspective it is of great interest to correlate the tissue mechanical response to the histological structure of the vocal fold ECM. For the human vocal fold age- and gender-dependent variations in collagen and elastin concentrations and their spatial distributions in the lamina propria were reported by Hammond *et al.*[23,24] Collagen concentration in the most superficial 40% of the lamina propria (cover) was observed to increase with age for males, but was found to be rather constant with age

for females. On the other hand, elastin concentration in the vocal fold cover was observed to increase with age for both males and females. When comparing the collagen content between geriatric males and females of 65 to 82 years of age, it was found that the collagen concentration in the vocal fold cover of males was considerably larger than in samples of females, whereas for elastin it was smaller in geriatric males than in females.[23,24] In the present study, for the male specimens it was possible to correlate the tissue mechanical response as expressed by the parameters of the constitutive model to their age. Consistent with the age-related increase in collagen and elastin concentrations it was observed that the tissue stiffness-characterizing parameters, μ_A and α, generally increase in magnitude with increasing age. These changes in tissue properties are well represented in the predicted fundamental frequency response, i.e., the increase of the predicted F_0 is correlated with the increase in the magnitude of the stiffness characterizing parameters.

In addition, the response of the male and the female specimens was compared with respect to the stiffness of the time-independent equilibrium response in network (A), i.e., $\bar{\mu}_A$. The present data suggest that stiffness of the male vocal fold cover is approximately twice that of the female, consistent with a higher collagen concentration (and perhaps lower elastin concentration) in geriatric males. Due to the small number of samples available within each gender group, however, this finding needs to be considered as preliminary, even though results of the t test indicated the difference was statistically significant. The parameter α determines the degree of nonlinearity of the tissue hyperelastic response, and its value qualitatively depends on the process of the recruitment of collagen and elastin fibers into the loading direction.[41] The lack of gender-related difference in $\bar{\alpha}$ could imply that the recruitment of collagen and elastin fibers in the vocal fold ECM to withstand tensile deformation is apparently gender unspecific. This was surprising, given that the male vocal fold cover has more collagen and less elastin than the female in geriatric specimens.[23,24] Further studies with additional specimens are needed to resolve this apparent discrepancy. In considering the consequent gender differences in F_0, gender-related magnitudes of vocal fold length changes are relevant. Adopting the finding that $\bar{\alpha}(\text{male}) \approx \bar{\alpha}(\text{female})$, $\bar{L} = 19.45$ mm (males, $Y > 50$) and $\bar{L} = 14.42$ mm (females), equal tissue densities for male and female, and a stretch of $\lambda_u = 1.2$, Eq. (15) predicts similar fundamental frequencies for the two genders, $\bar{F}_0(\text{male}) = 106$ Hz and $\bar{F}_0(\text{female}) = 104$ Hz. This finding suggests that the effect of longer vocal folds in our male specimens, which should lower F_0, is being compensated for by their higher tissue stiffness, which raises F_0. The higher F_0 values for females as documented from speech data could thus be the result of higher levels of stretch during vocal fold posturing.

The parameters characterizing the time-dependent response, μ_B and Z, decrease in magnitude with age. These parameter changes suggested that younger males may possess a compliant equilibrium network (A) and a relatively high stiffness in the elastic response of network (B). In such a situation, most of the stress would be carried by the network (B). Since the viscosity scaling constant Z of the inelastic flow rule was also large for the younger males, the resulting amount of inelastic deformation would be significant and lead to a pronounced hysteretic response. Older males may possess a stiffer equilibrium network (A) and a relatively compliant elastic response in network (B). In this case a reduced level of stress would be present in the time-dependent network (B). Since the viscosity scaling constant Z of the inelastic flow rule was small, the resulting amount of inelastic, time-dependent deformation would be reduced. This trend is represented in the increase in the value of the predicted time constants with age (Fig. 9).

VII. CONCLUSION

The mechanical behavior of the human vocal fold cover under large-strain uniaxial tensile loading was quantified at a frequency of 1 Hz, and data for six male and five female specimens were obtained. While this data set is limited and the findings should be considered preliminary, several conclusion could nonetheless be drawn:

(1) The stress-strain response of the vocal fold cover was found to be highly nonlinear and hysteretic, consistent with previous findings on vocal fold elasticity. Age and gender of the specimens appeared to significantly impact both the elastic and the time-dependent tensile deformation behavior. A constitutive model approach comprised of a hyperelastic (large-strain), equilibrium network in parallel with an inelastic (viscoplastic), time-dependent network was formulated to simulate the empirical data. It was found that the model was capable of describing the tissue behavior with reasonably high confidence, except for high levels of vocal fold elongation.

(2) Trends found in the age and gender dependence of the tissue mechanical behavior are in good agreement with previously reported age- and gender-related differences of collagen and elastin concentrations in the vocal fold cover. Based on the constitutive parameters it was found that for males the stiffness of the time-independent response increases with age and that the viscous hysteretic response of the tissue decreases with age. When comparing the gender groups, geriatric male specimens appeared to be nearly twice as stiff as the geriatric females. No significant difference between gender was found in the time-dependent response.

(3) Predictions of the fundamental frequency showed an increase in this parameter with age for males, in agreement with published life-span changes of F_0. The predicted characteristic time constants for F_0 under stretch transients are in the range of those from speech data. The magnitude of the time constants tends to increase with age.

(4) Considering equal magnitude of stretch, similar fundamental frequencies between geriatric males and females are predicted as the effect of shorter vocal fold length in females is compensated for by the reduced

tissue stiffness. No significant difference in the time-dependent response between males and females was found.

This study has focused on only the passive biomechanical response of the vocal fold cover, without including the important contributions of the vocal ligament in tensile stress regulation, and the active contributions of the intrinsic laryngeal muscles. Also, the size of our empirical data pool was limited. Nevertheless, it was demonstrated that some of the empirical life-span changes of speaking F_0 could be explained by the proposed constitutive model reasonably well. The present findings suggest that our modeling approach may provide a promising framework for characterizing the contributions of the passive tissue mechanical response to the process of F_0 regulation. Future investigations should include a larger number of specimens, a constitutive model including an extended formulation of inelastic deformation, and considerations regarding the contributions of the vocal ligament to fundamental frequency.

ACKNOWLEDGMENTS

This work was supported by the National Institute on Deafness and Other Communication Disorders, NIH Grant No. R01 DC006101. The authors thank Neeraj Tirunagari and Min Fu of UT Southwestern Medical Center for their assistance in empirical measurements of vocal fold tissue elasticity.

APPENDIX

The procedure followed in the parameter identification follows Ref. 18 but incorporates modifications for the current choice of the hyperelastic response. Initially, stresses in network (B) are set to zero (fully relaxing the network) by assigning a large viscosity scaling constant $Z(Z>1.0)$. Then, the hyperelastic equilibrium response provided by network (A) becomes the model response, characterized by the parameters μ_A and α. The values of μ_A and α are determined through a least-square optimization process following the Levenberg-Marquardt method to best fit the experimental equilibrium response estimated from the empirical stress-stretch curves as the midpoint values between the loading and unloading portions of the hysteresis loop. Subsequently, the initial shear modulus of network (B) is determined from the tangent stiffness E_t at load reversal where network (B) responds through its hyperelastic deformation only. The tangent modulus at the onset of unloading E_t is the sum of the elastic moduli of the two networks:

$$E_t = 3\mu_B + \frac{2\mu_A}{\alpha}\left[(\alpha-1)\lambda_{rev}^{\alpha-2} + \left(\frac{\alpha}{2}+1\right)\lambda_{u,rev}^{-\alpha/2-2}\right]. \quad \text{(A1)}$$

The viscosity scaling constant Z is then obtained such that the simulated loading-unloading curves expand gradually from the equilibrium curve to fit the experimental data points of the stress-stretch hysteresis loop. The stress power parameter m is chosen such that the predicted response fits the initial slope of the experimental loading curves. For all but one of the present specimens of the present investigation the stretch power parameter is set to $c=-1.0$. With this choice the model allows one to predict the decay in hysteretic loss with an increase in $\lambda_{u,rev}$ as observed previously.[4]

[1]E. J. Hunter, I. R. Titze, and F. Alipour, "A three-dimensional model of vocal fold abduction/adduction," J. Acoust. Soc. Am. **115**, 1747–1759 (2004).
[2]I. R. Titze, J. G. Švec, and P. S. Popolo, "Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues," J. Speech Lang. Hear. Res. **46**, 919–932 (2003).
[3]I. R. Titze, "On the relation between subglottal pressure and fundamental frequency in phonation," J. Acoust. Soc. Am. **85**, 901–906 (1989).
[4]R. W. Chan and T. Siegmund, "Vocal fold tissue failure: preliminary data and constitutive modeling," J. Biomech. Eng. **126**, 466–474 (2004).
[5]F. Alipour, D. A. Berry, and I. R. Titze, "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012 (2000).
[6]H. E. Gunter, "A mechanical model of vocal-fold collision with high spatial and temporal resolution," J. Acoust. Soc. Am. **113**, 994–1000 (2003).
[7]F. Alipour-Haghighi and I. R. Titze, "Elastic models of vocal fold tissues," J. Acoust. Soc. Am. **90**, 1326–1331 (1991).
[8]Y. B. Min, I. R. Titze, and F. Alipour-Haghighi, "Stress-strain response of the human vocal ligament," Ann. Otol. Rhinol. Laryngol. **104**, 563–569 (1995).
[9]T. Y. Hsiao, C. L. Wang, C. N. Chen, F. J. Hsieh, and Y. W. Shau, "Elasticity of human vocal folds measured *in vivo* using color Doppler imaging," Ultrasound Med. Biol. **28**, 1145–1152 (2002).
[10]A. P. Lobo and M. O'Malley, "A nonlinear finite-element model of the vocal fold," J. Acoust. Soc. Am. **99**, 2473–2500 (1996).
[11]R. W. Chan and I. R. Titze, "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021 (1999).
[12]R. W. Chan and I. R. Titze, "Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling," J. Acoust. Soc. Am. **107**, 565–580 (2000).
[13]M. S. Green and A. V. Tobolsky, "A new approach to the theory of relaxing polymeric media," J. Chem. Phys. **14**, 80–92 (1946).
[14]R. W. Ogden, "Larger deformation isotropic elasticity – on the correlation of theory and experiment for incompressible rubberlike solids," Proc. R. Soc. London, Ser. A **326**, 565–584 (1972).
[15]K. Miller and K. Chinzei, "Mechanical properties of brain tissue in tension," J. Biomech. **35**, 483–490 (2002).
[16]D. F. Meaney, "Relationship between structural modeling and hyperelastic material behavior: application to CNS white matter," Biomech. Model. Mechanobiol. **1**, 279–293 (2003).
[17]J. S. Bergström and M. C. Boyce, "Constitutive modeling of the large strain time-dependent behavior of elastomers," J. Mech. Phys. Solids **46**, 931–954 (1998).
[18]J. S. Bergström and M. C. Boyce, "Constitutive modeling of the time-dependent and cyclic loading of elastomers and application to soft biological tissues," Mech. Mater. **33**, 523–530 (2001).
[19]J. S. Bergström and M. C. Boyce, "Large strain time-dependent behavior of filled elastomers," Mech. Mater. **32**, 627–644 (2000).
[20]C. P. Buckley and D. C. Jones, "Glass-rubber constitutive model for amorphous polymers near the glass transition," Polymer **36**(17), 3301–3312 (1995).
[21]M. B. Rubin and S. R. Bodner, "A three-dimensional nonlinear model for dissipative response of soft tissue," Int. J. Solids Struct. **39**, 5081–5099 (2002).
[22]J. E. Bischoff, E. M. Arruda, and K. Grosh, "A rheological network model for the continuum anisotropic and viscoelastic behavior of soft tissue," Biomech. Model. Mechanobiol. **3**, 56–65 (2004).
[23]T. H. Hammond, S. D. Gray, and J. E. Butler, "Age- and gender-related collagen distribution in human vocal folds," Ann. Otol. Rhinol. Laryngol. **109**, 913–920 (2000).
[24]T. H. Hammond, S. D. Gray, J. Butler, R. Zhou, and E. H. Hammond, "A study of age and gender related elastin distribution changes in human vocal folds," Otolaryngol.-Head Neck Surg. **119**, 314–322 (1998).
[25]I. R. Titze and E. J. Hunter, "Normal vibration frequencies of the vocal ligament," J. Acoust. Soc. Am. **115**, 2264–2269 (2004).
[26]S. D. Gray, I. R. Titze, R. Chan, and T. H. Hammond, "Vocal fold proteoglycans and their influence on biomechanics," Laryngoscope **109**, 845–854 (1999).
[27]S. D. Gray, I. R. Titze, F. Alipour, and T. H. Hammond, "Biomechanical

and histological observations of vocal fold fibrous proteins," Ann. Otol. Rhinol. Laryngol. **109**, 77–85 (2000).
[28]MATLAB, The MathWorks, Inc., Natick, MA (Version 7.0).
[29]A. L. Perlman, "A technique for measuring the elastic properties of vocal fold tissue," Ph.D. dissertation, The University of Iowa, Iowa City, IA, 1985.
[30]S. D. Gray, K. J. Chan, and B. Turner, "Dissection plane of the human vocal fold lamina propria and elastin fibre concentration," Acta Oto-Laryngol. **120**, 87–91 (2000).
[31]J. Ohala and W. Ewan, "Speed of pitch change," J. Acoust. Soc. Am. **53**, 345A (1973).
[32]J. Sundberg, "Maximum speed of pitch changes in singers and untrained subjects," J. Phonetics **7**, 71–79 (1979).
[33]I. R. Titze, J. J. Jiang, and E. Lin, "The dynamics of length change in canine vocal folds," J. Voice **11**, 267–276 (1997).
[34]S. J. Lai-Fook and R. E. Hyatt, "Effects of age on elastic moduli of human lungs," J. Appl. Physiol. **89**, 163–168 (2000).
[35]J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, *Applied Linear Statistical Models*, 4th ed. (McGraw-Hill, New York, 1996).
[36]H. Hollien, "Vocal pitch variation related to changes in vocal fold length," J. Speech Hear. Res. **3**, 150–156 (1960).
[37]W. Brown, R. Morris, Jr., H. Hollien, and E. Howell, "Speaking fundamental frequency characteristics as a function of age and professional singing," J. Voice **5**, 310–315 (1991).
[38]M. I. P. Krook, "Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis," Folia Phoniatr. **40**, 82–90 (1988).
[39]I. R. Titze, *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[40]M. Döllinger, D. A. Berry, and G. S. Berke, "Medial surface dynamics of an in vivo canine vocal fold during phonation," J. Acoust. Soc. Am. **117**, 3174–3183 (2005).
[41]P. Fratzl, K. Misof, I. Zizak, G. Rapp, H. Amenitsch, and S. Bernstorff, "Fibrillar structure and mechanical properties of collagen," J. Struct. Biol. **122**, 119–122 (1997).

Characterizing glottal jet turbulence

Fariborz Alipour[a)]
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

Ronald C. Scherer
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242, and Department of Otolaryngology—Head and Neck Surgery, University of Cincinnati Medical Center, Cincinnati, Ohio 45267-0528

(Received 13 January 2005; revised 10 November 2005; accepted 22 November 2005)

Air pressure associated with airflow from the lungs drives the vocal folds into oscillation and allows the air to exit the glottis as a turbulent jet, even though laminar flow may enter the glottis from the trachea. The separation of the turbulence from the deterministic portion of the glottal jet was investigated in the excised canine larynx model. The present study is methodological in that the main goal was to examine three methods of obtaining reasonable representations of both the deterministic signal and the residual turbulence portion: (a) smoothing, (b) wavelet denoising, and (c) ensemble averaging. Ensemble averaging resulted in a deterministic signal that disregarded gross cyclic alterations while exaggerating the turbulence intensity. Wavelet denoising can perform an excellent analysis and synthesis of the glottal velocity, but was problematic in determining which levels of analysis to choose to represent both the deterministic and turbulence appropriately. Smoothing appeared to be the most appropriate for phonation velocities because it preserved gross cyclic variations important to perturbations and modulations, while extracting turbulence at what appears to be reasonable levels. [DOI: 10.1121/1.2151809]

PACS number(s): 43.70.Aj, 43.70.-h [AL] Pages: 1063–1073

I. INTRODUCTION

Glottal flow during phonation exits the glottis as a pulsatile jet that has both a repetitive quasiperiodic magnitude and shape (called the *deterministic signal*), as well as a more random *turbulence* ripple (Alipour *et al.*, 1995). The source of turbulence in real glottal flow may arise from jet flow through an incompletely closed posterior glottis, as well as from pulsatile jet behavior of the air moving through the membranous glottis (Hillman *et al.*, 1983). Studies have suggested that the glottal jet between the vocal folds is typically laminar near the glottal exit, with transitions to turbulence within a short distance of the glottis (Alipour *et al.*, 1995; Shinwari *et al.*, 2003; Hofmans *et al.*, 2003; Kucinschi, 2004; Kucinschi *et al.*, in review).

It is generally, and reasonably, assumed that turbulence in the laryngeal flow acoustically excites the vocal tract, adding to the quality of the perceived sound. This turbulence most likely is the primary differentiator for breathy quality variations. This has led to studies dealing with acoustic correlates of breathiness (e.g., Hillenbrand and Houde, 1996; de Krom, 1995; Wolfe and Martin, 1997), motion of the glottis related to voice qualities (Cooke *et al.*, 1997), glottal flow correlates of breathiness (e.g., Alku and Vilkman, 1996), models related to breathiness (e.g., Childers and Lee, 1991; Childers and Ahn, 1995), and auditory modeling connections to the perception of breathiness (Shrivastav and Sapienza, 2003). What is missing are studies dealing directly with measuring the turbulence of the glottal flow to determine its characteristics related to glottal flow phenomena, as well as to the perception of breathiness and voice quality control in general. First, however, methods need to be established to extract the turbulence in the glottal flow, and this is the subject of this paper.

The study of turbulence in physiological pulsatile flow such as blood flow in heart valves and stenoses can be performed in static transparent models using laser Doppler anemometry (Einav and Sokolov, 1993; Baldwin *et al.*, 1993; Bluestein and Einav, 1995). However, glottal flow is complicated by the complex dynamic geometry of the glottis and high rates of vocal fold oscillation and airflow through the glottis. The most suitable model for the purpose of studying glottal flow in detail may be the canine excised larynx model. This model provides a pulsatile flow with phonation, and the glottal exit flow is turbulent above the glottis (Berke *et al.*, 1989; Alipour *et al.*, 1995). It is important to be able to determine the turbulence that is present within the air flowing from the glottis during phonation, as well as to calculate a meaningful turbulence intensity that may have perceptual and acoustic importance.

The glottal flow is a pulsatile or modulated flow that is caused by the oscillation of the vocal folds under the driving influence of translaryngeal pressure. For phonation with a quality that is perceived to be normal (instead of highly rough, for example), the flow has a dominant (deterministic) structure that is most typically thought of as a modulated volume velocity with shape and size that are related to primary acoustic excitation of the vocal tract. This volume flow is an integration of the instantaneous particle velocities

[a)]Please address all correspondence to Fariborz Alipour, Ph.D., The University of Iowa, 334 WJSHC, Iowa City, Iowa 52242-1012. Telephone: (319) 335-8694; fax: (319) 335-8851; electronic mail: alipour@blue.weeg.uiowa.edu

across the glottis (Alipour *et al.*, 1995; Alipour and Scherer, 1995). A more basic study of laryngeal aerodynamics (and aeroacoustics) would be to examine the modulated velocity structure in more detail, namely, to attempt to elucidate and separate the structured velocity signal that cyclically repeats itself from the incoherent turbulence that arises due to glottal jet flow instabilities.

In this paper we discusse three different methods of analyzing the velocity signal of the air exiting the glottis into the deterministic (nearly repetitive) signal and the turbulence fluctuation velocities. These methods are *smoothing (filtering), wavelet denoising*, and *ensemble averaging*. Walburn *et al.* (1983) compared ensemble averaging with high pass digital filtering in animal and human aortic valves, and found that the absolute turbulence intensity from ensemble averaging was significantly higher than using digital filtering. The study here will also show similar results, suggesting that ensemble averaging leads to an inadequate matching of the velocity cycles, and exaggerates turbulence intensity. It is also noted that Sullivan *et al.* (1999) performed a similar experiment in a repeating cyclic system (spark ignition engines) by comparing three methods (quite similar to those used here) to obtain the mean and turbulent velocities, namely, ensemble averaging, cyclic averaging, and wavelet averaging. They found that the methods gave similar average velocity signals, but gave differences in the turbulence velocities, suggesting that those differences could result in subjectively alternative interpretations of the flow physics. The results of the study presented here suggest that the methods differ both in how well they follow individual cycles and in how similar are the turbulence velocity signals. It is clear that there are numerous methods that might be used to obtain the deterministic (repeating) flow structures and turbulence in velocity signals from the glottis, and this study compares three such methods that are also easily available.

It is stressed that a study of methods to adequately extract the deterministic and turbulence signals from glottal velocity signals is part of a broader practical issue of measuring (and predicting) the generation of aerodynamic noise in the larynx and vocal tract, and establishing turbulence "management and control" (Hussain, 1986), which in this case would relate to surgical, therapy, and training practices to improve laryngeal control, voice quality, and a patient's or client's communication skills.

II. METHODOLOGY

A. Excised experimental arrangement

The techniques used in this experiment were similar to those used in earlier studies (Alipour *et al.*, 1995), and will be briefly summarized here. Canine larynges were acquired from other medical laboratories at the University of Iowa and kept in saline solution prior to use. Figure 1 shows a mounted excised larynx with a hot-wire probe (Dantec P14 with right-angled prongs and sensor direction perpendicular to the flow axis) placed above the glottis centerline at a distance of approximately 10 mm. The attempt was to place the wire of the hot-wire probe as close to the glottal exit as possible without endangering the integrity of the probe via

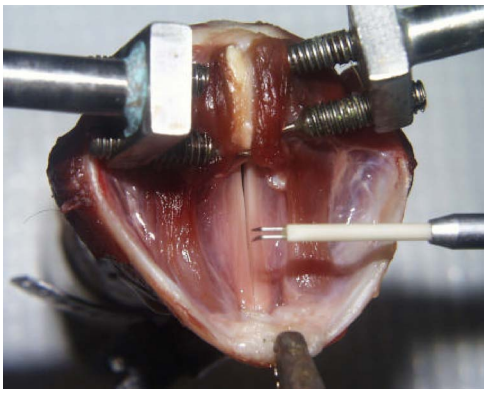


FIG. 1. (Color online) Mounted excised larynx with adduction prongs pushing arytenoids, hot-wire sensor positioned over the glottis midline.

contact from moving vocal fold tissue or from saline and mucosal splash. Another limitation was the necessity to clear the edge of the trimmed thyroid cartilage from the side, as shown in Fig. 1. The probe construction appeared to be such that it should have negligible effect on velocity measures. It is noted that the probe had a single wire and was not sensitive to the direction of the velocity coming toward it. The assumption was made, however, that because of the position of the wire above the centerline of the glottis, the primary axis of the velocities acquired was axial (vertical). No vocal tract was used in this experiment. Air from a pseudolung passed through a heater and humidifier (Concha Therm® unit, RCI Laboratories) to achieve 37 °C and 100% humidification. The use of sutures, micrometers, and shims controlled glottal length and adduction, although length was not a parameter that was varied in this study. Electrode plates from an electroglottograph (Synchrovoice) were placed on the thyroid laminae to obtain the EGG signal during phonation. The EGG signal waveform was used to extract fundamental frequency and jitter (see Titze *et al.*, 1987, for definitions), and to obtain cycle information for phase shift averaging.

The time-varying subglottal pressure was recorded using a pressure transducer (Microswitch 136PC01G1) mounted in the tracheal tube 10 cm below the vocal folds, while the mean subglottal pressure was monitored on a well-type manometer (Dwyer Model 1230-8) through a pressure tap across from the pressure transducer. The mean flow rate was monitored with an in-line flow meter (Gilmont rotameter model J197, ±3%). The velocity in the trachea was measured by using a straight miniature probe (Dantec P11) that was positioned in the center of the subglottal tubing (perpendicular to the flow) approximately 12 cm below the vocal folds. The hot-wire probes were calibrated using a Dwyer stainless steel Pitot tube (model 167-6) to correlate the output signals of the hot wire probe to their corresponding calibrated velocities. After collecting 15 to 20 data points that ranged from 0 to the maximum expected velocity (50 m/s), a fourth-order polynomial was fitted to the data points and coefficients were saved in a file for data calibration. Velocity measurement accuracy was estimated to be no worse than ±4% in the experimental range.

Analog signals from the hot-wire sensors, EGG, and

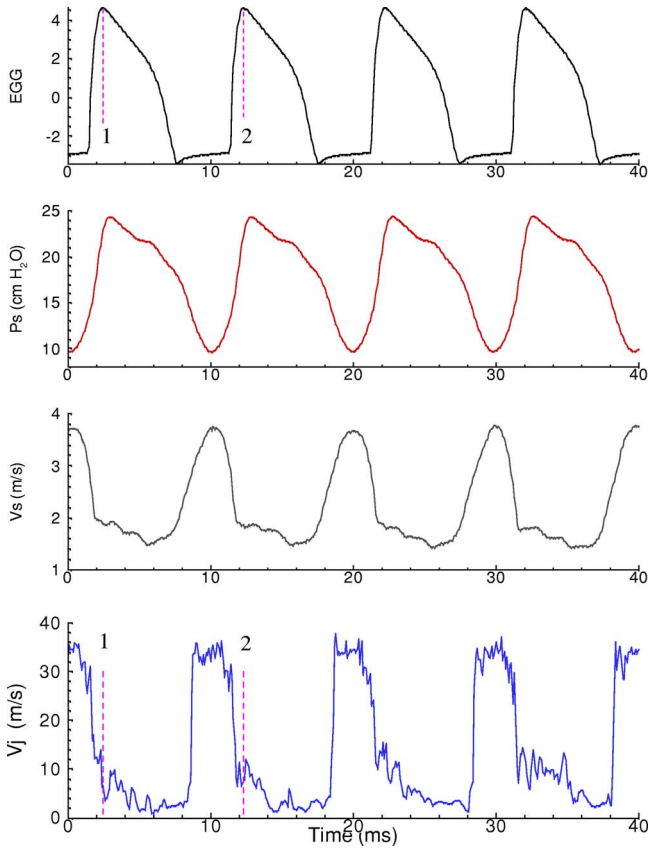


FIG. 2. (Color online) A few cycles of the typically recorded signals, including from top to bottom: Electroglottograph (EGG), subglottal pressure (Ps), tracheal velocity (Vs), and glottal exit jet velocity (Vj).

pressure transducer were monitored on a digital oscilloscope (Tektronix, TDS2014). Two systems were used at different times to simultaneously record the various signals for 1 to 5 s. One system included a Digital Sound Corporation A/D converter, sampling at 20 kHz per channel (antialiasing at 10 kHz) directly to computer files. The other system simultaneously recorded the signals onto a DAT recorder (Sony, PC-108M) and later digitized at 10 kHz per channel (antialiasing at 5 kHz) using a DATAQ A/D converter. These signals were converted to physical quantities in MATLAB and used for the turbulence analysis.

Figure 2 shows typically recorded data, including from top to bottom the following signals: EGG, subglottal pressure (Ps), tracheal velocity (Vs), and glottal exit jet velocity (Vj). These data correspond to the oscillations of an excised larynx at 101 Hz with a mean flow rate of 440 ml/s and an extremely low jitter of 0.09% (for 200 cycles). The variations in the glottal velocity signal suggest the existence of turbulence in the glottal jet, and the lack of such variations in the tracheal velocity signal (laminar flow). The top view of the larynx was recorded onto videotape via a camera. The tissue was illuminated by a stroboscope (Monarch Instruments, Phaser Strobe PB) triggered manually or externally by the EGG signal. For selected cycles, a frame of maximum opening was captured using a frame grabber (Integral Technologies, FlashBus MV, 640×480 resolution). A plastic sheet with a millimeter grid (5×10) was placed on the vocal folds spanning the glottis and was video recorded at the beginning of every case that might include vocal fold positioning for later calibration of the images. Using the MATLAB Image Processing Toolbox and custom-made software, vocal fold length (L_0), maximum glottal area (A_g), maximum glottal width (G_w), and the glottal perimeter (G_p) were measured. In this procedure, the edge of the glottis in full screen image was traced with the mouse to generate a polygon region of interest and then major properties of this region were calculated and calibrated against a previously recorded grid image.

B. Turbulence extraction techniques

In general, turbulent flows, considered being major mechanisms of mass, momentum, and energy transfer, contain random fluctuations. A precise definition of turbulence is difficult. We can recognize turbulent flow from some of its characteristics such as high Reynolds number, irregularity, high diffusivity, dissipation rate, three-dimensionality, and vorticity fluctuations. Turbulence is initiated by flow instabilities. A glottal jet is a shear flow that becomes very unstable at Reynolds number as low as 10 (White, 1974). A detailed description and analysis of turbulence structures are beyond the scope of this paper. In the interest of flow analysis, it is important to separate these random fluctuations.

The instantaneous velocity in the glottal jet can be written as the sum of the coherent (deterministic) and incoherent (random fluctuations) components as

$$V = V_d + u'. \tag{1}$$

The deterministic component (V_d) can also be written as the sum of an oscillatory and a stationary component as

$$V_d = U_{\mathrm{os}} + \bar{U}. \tag{2}$$

In steady flows the oscillatory component (U_{os}) drops out and by averaging instantaneous velocity over long periods of time, one can easily find the mean or time-averaged velocity ($\bar{U}$), and by subtraction, an estimate of the turbulence. However, in pulsatile glottal flow, due to the existence of the oscillatory component, simple time averaging does not work and other ways of separating turbulence are needed. Three of these techniques of obtaining the deterministic signal are described below. Once the deterministic velocity is obtained, the subtraction of that velocity from the instantaneous velocity is the turbulent fluctuation velocity.

1. Smoothing (filtering)

In this method, a nonrecursive digital low-pass filter (Hamming, 1973) was used to smooth the jet velocity signal. The filter used seven consecutive sample points and is based on a moving average of the sample points with the following form:

$$g_k = (f_{k-3} + 6f_{k-2} + 15f_{k-1} + 20f_k + 15f_{k+1} + 6f_{k+2} + f_{k+3})/64, \tag{3}$$

where f_k refers to the raw velocity at point (k) and g_k refers

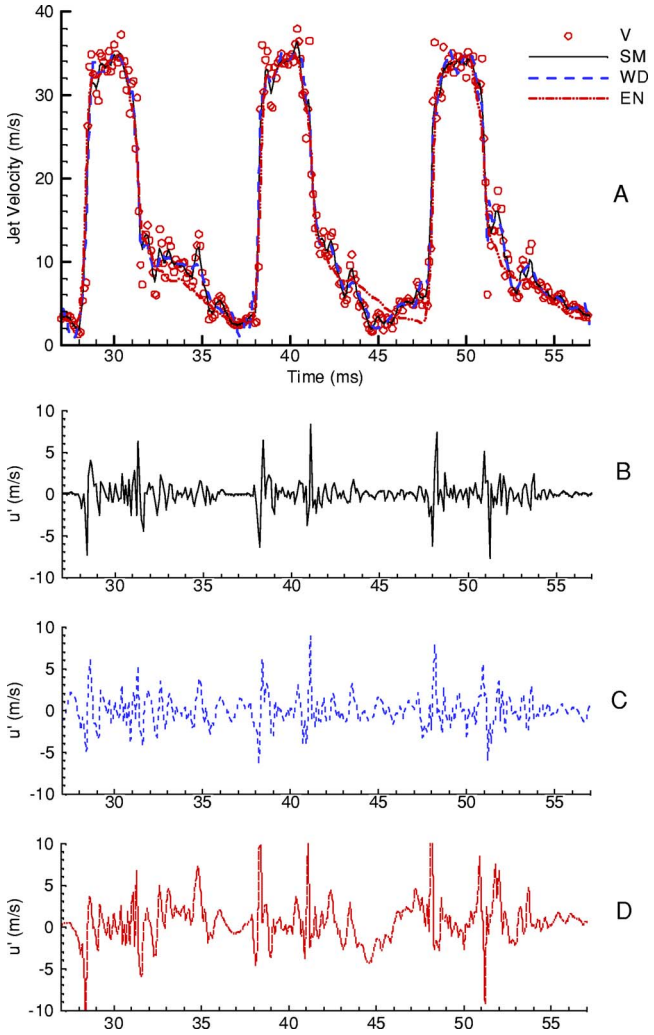


FIG. 3. (Color online) Turbulence averaging for a low-jitter larynx (0.05%). In the top graph (a) data points of instantaneous velocity (circles) and averaged signals are plotted together for three cycles. These are the smoothed velocity (solid line), wavelet denoised (dashed), and ensemble-averaged (dash-dot-dot) velocities. The corresponding turbulent velocity fluctuations are graphed below them [(b), (c), (d)].

to the filtered velocity at that same location. This smoothing process averages each sample point of the velocity signal with its six neighboring points. The larger weights for the nearest neighbors (20, 15, and 6) preserve most of the desired features in the glottal signal such as jitter and shimmer. The transfer function for this filtering process can be obtained by substituting $f_k = e^{i\omega k}$ as

$$H(\omega) = \frac{g_k}{f_k} = \frac{5}{16} + \frac{15}{32}\cos(\omega) + \frac{3}{16}\cos(2\omega) + \frac{1}{32}\cos(3\omega). \tag{4}$$

This function attenuates the input signal to 45.7% at $\omega = 1$ and to 2.5% at $\omega = 2$, where frequencies (f) are normalized to sampling frequency f_s as

$$\omega = 2\pi f / f_s. \tag{5}$$

Figure 3(a) shows an example of signal averaging with this smoothing method. In the top graph, data points of instantaneous velocity (circles) and the smoothed signal (solid line) are plotted together for three cycles. On the second graph (b), the difference between the raw velocity values and the smoothed signal, that is, the *turbulent fluctuation velocity* is shown. Sharp spikes can be seen at locations of abrupt rise and fall in the smooth signal.

2. Wavelet denoising

Wavelet denoising was the second technique used to separate the turbulence from the deterministic structure of the velocity signal. In this method, unlike the Fourier transform, which decomposes the signal into periodic components, wavelet denoising uses a template-matching transform with different resolutions or scales to approximate the signal (Resnikoff and Wells, 1998). In other words, linear combinations of wavelet functions such as Daubechies, Coifman, and others, can be used to approximate different portions of the signal. This feature arises from the time-limited (localized) properties of wavelet functions, which allow them to represent any portion of a signal efficiently, even when there is discontinuity. Wavelets decompose any signal into multiresolution components, a coarse resolution and multiple levels of fine resolution. For each level, a series of coefficients are calculated that will be used in the reconstruction of the signal. When a wavelet transform is applied to the velocity signal, the properties of the signal can be emphasized or deemphasized by manipulating these coefficients. There are choices of a wavelet function and its scale, and the number of levels of the wavelet transform. The turbulence in the glottal jet velocity was separated with various resolution levels and choices of wavelet functions using the "denoising function" of the MATLAB Wavelet Toolbox. Once a wavelet function is selected, the velocity signal is denoised in MATLAB with the default soft threshold parameter (THR), which is calculated from

$$\mathrm{THR} = S\sqrt{2\log(n)}, \tag{6}$$

where (S) is the estimate of the noise level and (n) is the sample size. By discarding those coefficients smaller than THR, the reconstructed signal will be denoised. Thus, a larger value of THR may give a cleaner deterministic signal, but we may lose some important information in the process. While the THR was not controlled as a parameter, the choices of wavelet function and the level of resolution and sensitivity of results were, and are compared in Table I and discussed in Sec. III. Here the choice of "db3," "sym3," and "coif3" gave similar results and the level 3 gave a denoised velocity with higher cross-correlations with the original signal. Figures 3(a) and 3(c) show an example of a wavelet denoised velocity (dashed line) and turbulent fluctuations.

3. Ensemble averaging

For periodic or quasiperiodic flows with period (T), the mean value of the velocity at some particular phase point in a cycle can be obtained by sampling over many cycles as

TABLE I. A comparison of the turbulence-averaging methods. Column 4 indicates the model with the value of "SMZ" for smoothing and "ENS" for the ensemble averaging; otherwise it is for wavelet denoising. Column 5 (Par) is the number of cycles for smoothing or ensemble averaging; otherwise it is the level of the wavelet denoising. Tint is the turbulent intensity, U_{bar} is the time average of velocity fluctuations, and CCOR is the cross-correlation between denoised and original velocity signals.

Larynx No.	Fo	Jitter from EGG	Model	Par	Tint	U_{bar}	CCOR
2	101.1	0.05	SMZ	1	1.51	8.89E-05	0.9968
			ENS	200	2.58	3.37E-02	0.99077
			db3	1	1.14	8.59E-05	0.99814
			sym3	1	1.14	8.59E-05	0.99814
			coif3	1	1.14	1.97E-05	0.99817
			coif3	2	1.55	−2.06E-05	0.99657
			coif3	3	1.89	−2.55E-04	0.99493
			coif3	4	2.11	3.69E-05	0.99365
			coif3	5	2.26	9.24E-04	0.99289
3	145.5	5.89	SMZ	1	0.87	6.29E-06	0.99785
			ENS	400	3.65	1.40E-01	0.96082
			db3	1	0.6	−8.21E-06	0.99893
			sym3	1	0.6	−8.21E-06	0.99893
			coif3	1	0.58	4.53E-06	0.99901
			coif3	2	0.87	−1.24E-06	0.99775
			coif3	3	1.07	8.56E-05	0.9966
			coif3	4	1.23	2.04E-04	0.9956
			coif3	5	1.31	6.81E-05	0.99523
4	136.8	12.35	SMZ	1	0.89	3.85E-05	0.99872
			ENS	400	4.91	1.43E-01	0.96705
			db3	1	0.61	3.16E-05	0.99939
			sym3	1	0.61	3.16E-05	0.99939
			coif3	1	0.56	9.68E-06	0.99949
			coif3	2	1.08	9.01E-05	0.99808
			coif3	3	1.58	8.45E-05	0.99584
			coif3	4	1.86	2.35E-04	0.99426
			coif3	5	2.01	−1.33E-04	0.99337

$$\langle V(x,t)\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N} V(x,t+iT). \tag{7}$$

In this process, which is called ensemble averaging (also called phase averaging), the EGG signal was used as the trigger signal to obtain glottal cycle information (period T, and phase point locations). Every instantaneous velocity measurement on consecutive cycles corresponding to a phase point on the trigger signal was averaged over 100 to 500 cycles (Walburn *et al.*, 1983; Sullivan *et al.*, 1999). The phase points were selected at the peak of the EGG signal to preserve the pitch variations in the phase-averaged velocity signal. The cycle lengths differed across the cycles. In order to make each cycle equal in length, the maximum-length cycle was determined, and the shorter cycles were padded with the appropriate end point section of the longest cycle rather than with zeros (in order to reduce the bias that zeros would create, including unrealistic transients near the end of the processed cycles). In Fig. 2, two locations on the EGG signal and jet velocity with corresponding phase points of (1) and (2) are identified with vertical dashed lines. These two locations are the beginning and the end of the phase points on that cycle. The averaging process required determination of all such peak locations on the EGG signal, adding the magnitude values at all of the corresponding phase points, and then averaging them. A low-pass filtering of the EGG signal at 500 Hz helped to stabilize the process by removing unwanted higher-frequency noise ripple on the EGG signal. The ensemble averaging provided the deterministic component of the oscillatory flow (Alipour *et al.*, 1995). The larger the number of cycles included in the averaging, the smoother the signal that is obtained. The ensemble-averaged velocity signal represents an ideal deterministic velocity signal. When this signal is subtracted from the instantaneous velocity, the turbulent component is obtained. Figures 3(a) and 3(d) show the ensemble-averaged signals (dash-dot-dot line). Here, 200 cycles of the velocity signal were processed and the resulting averaged signal shows *identical* cycle shapes. The ensemble averaging technique is reminiscent of obtaining the harmonics-to-noise ratio suggested by Yumoto *et al.* (1982), whereby the average (deterministic) cycle was obtained, subtraction of that cycle from each of the original cycles gave the "noise" signal, and the ratio of the energy of the deterministic and noise signals gave the desired ratio.

C. Absolute turbulence intensity

The three methods discussed above provided the deterministic velocity and turbulence fluctuations for the same

velocity signal. To compare these methods, the *absolute turbulence intensity* was calculated, and is defined as

$$\alpha = \sqrt{\overline{(u')^2}}, \tag{8}$$

which is the root means square (RMS) of the fluctuation velocity (Walburn *et al.*, 1983). Also, since u' is a random quantity, its time average theoretically should be zero. These two quantities may be used as criteria for comparing these methods. Table I shows the comparative values for these quantities with various parameters for three larynges. For each larynx, fundamental frequency and jitter values are given in columns 2 and 3. Column 4 indicates the turbulence-averaging models with the value of "SMZ" for smoothing, "ENS" for the ensemble averaging, and others, namely "db3," "sym3," and "coif3," for wavelet denoising. Column 5 shows the parameter of the model, which is the number of cycles for ensemble averaging and the level for wavelet denoising. Columns 6 and 7 are the turbulent intensity and the time average of the turbulent velocity fluctuations. The last column represents the mean cross-correlation of the averaged signal and the original signal.

D. Direct and derived measures

In order to indicate the range of the phonatory conditions in this experiment, a number of nondimensional terms were used. They required knowledge of the air pressures, airflows, and glottal geometry. The recorded images were reviewed and those with higher quality and clarity were digitized on a PC computer using a video capture card. Using slow motion and frame-by-frame inspection of the images, the frames with the maximum glottal opening were captured and stored. These images were analyzed for glottal geometric characteristics using MATLAB software. The length and area were calculated and calibrated against a previously recorded millimeter grid that had been placed on the top of the vocal folds.

The maximum glottal area (A_g) and the corresponding measured glottal perimeter (G_p) were used to obtain the hydraulic diameter ($D_h = 4A_g/G_p$) of the glottis during oscillation. The mean Reynolds number was calculated based on this length scale ($\mathrm{Re} = U_m D_h/\nu$), where U_m is the mean velocity through the glottis, and ν is the kinematic viscosity. The mean glottal velocity, U_m, was calculated ($\pm 5\%$) by dividing the mean volume flow rate obtained from the rotameter by the maximum glottal area. The Strouhal number was calculated based on these parameters and the frequency of vocal fold oscillation ($\mathrm{St} = fD_h/U_m$).

III. RESULTS AND DISCUSSION

A. Basic relations for derived measures

The ranges of values for the aerodynamic and derived terms are important as a basis for interpretation of the turbulence results presented later. In the experiments reported here, six excised larynges were used; subglottal pressures ranged from 7.4 to 29.2 cm H_2O, and flows ranged from about 100 to 1000 ml/s, thus covering the approximate full range expected in typical human phonation. The hydraulic diameter of the glottis ranged from about 0.06 to 0.47 cm, also a relatively wide range expected for human phonation. The flow and hydraulic diameter values were related approximately monotonically to each other for each condition (Fig. 4). A regression line with the slope of 5.323×10^{-4} cm/(ml/s) is drawn through the data points for four larynges.

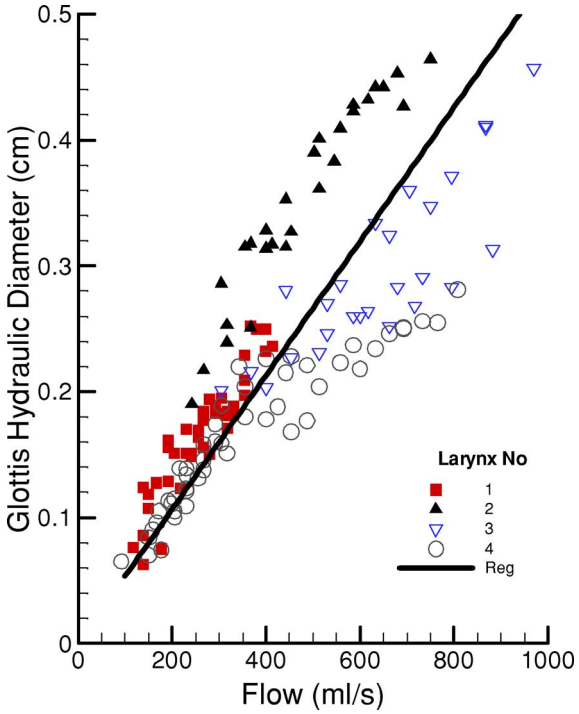


FIG. 4. (Color online) Glottis hydraulic diameter as a function of flow rate for four excised larynges.

Figure 5 indicates the range of values for the Reynolds and Strouhal numbers used in the experiments for six canine larynges. The Reynolds number ranged from about 1200 to 6800. The Strouhal number ranged from about 0.002 to 0.03. These ranges are consistent with those used in Alipour *et al.* (1995). It is expected that the function of the human larynx with similar dimension to the canine (15 mm average vocal fold length in both) may result in comparable ranges of Reynolds and Strouhal numbers. These results for the general descriptors of the phonation cases of this study suggest that the turbulence discussions below should be representative of that which occurs in human phonation. A regression line with the slope of 3.1463×10^{-6} is drawn through the data points for the six larynges.

B. Turbulence extraction comparisons

1. Three cases based on jitter level

a. Case 1: low jitter ($J = 0.05\%$). Figure 3 discussed above gives the jet velocity values (circle data points not connected by lines), the turbulence-removed velocity, and the turbulent fluctuation velocities from each method. The *smoothing filter* technique (solid line) results in a fairly robust fit to the scatter of data throughout the signal, despite the obvious cyclic variations from cycle to cycle, with relatively strong peaks during the steep gradient sides (left and right) on the jet velocity signal (upper trace) and a turbulent intensity value of 0.656 m/s. The *wavelet denoising* ap-

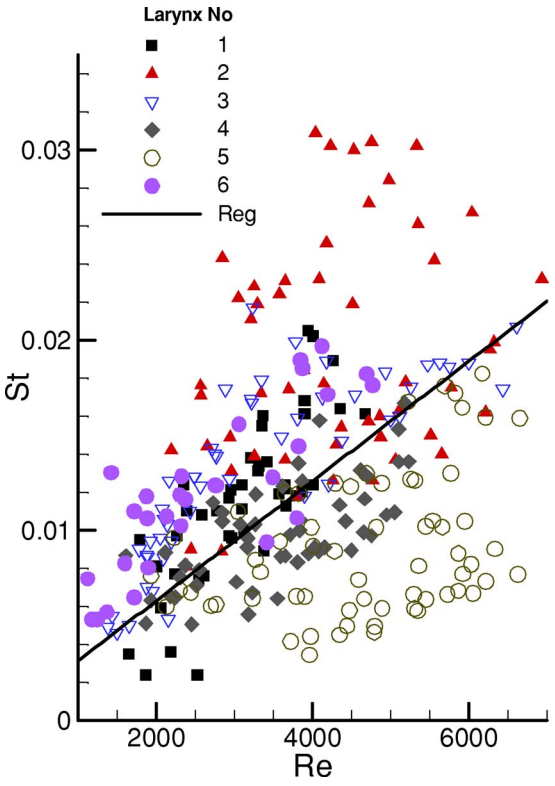


FIG. 5. (Color online) Aerodynamic ranges of the excised canine larynges. The graph shows the ranges for the Reynolds and Strouhal numbers for all excised larynges used in this study.

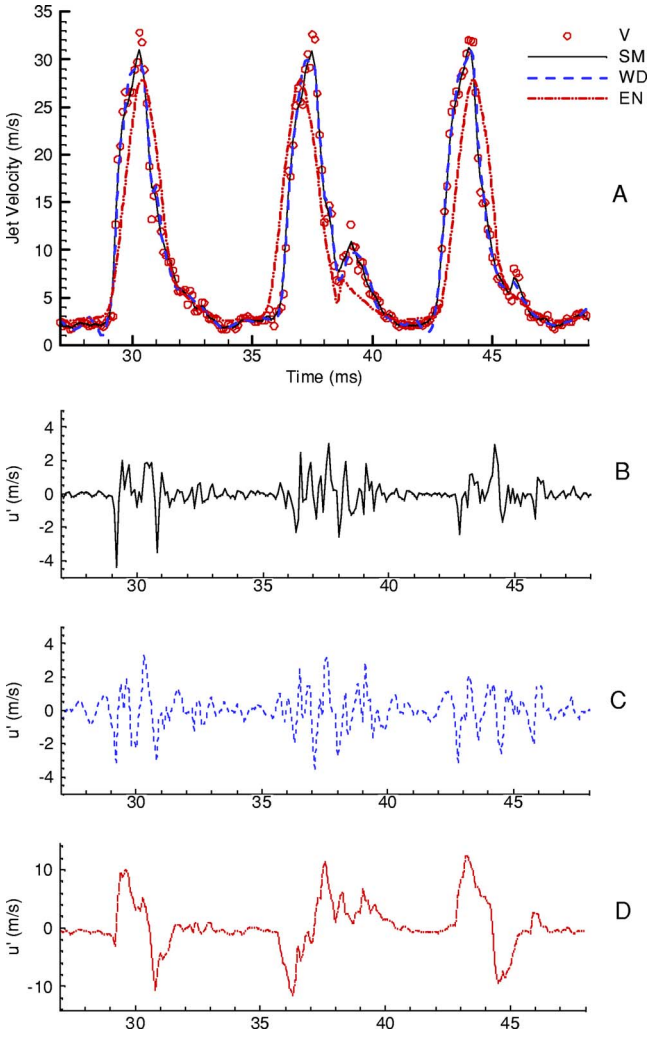


FIG. 6. (Color online) Turbulence averaging for a high-jitter larynx (5.89%). The descriptions of the signals are the same as Fig. 3.

proach using level 3 here (dashed line) creates a closer match to the original signal, with a subsequent reduction in the u' peak values compared to the other methods and turbulence intensity of 0.324 m/s. It may appear, however, that too much turbulence-appearing variation remains in the denoised signal. For the *ensemble averaging* (dash-dot-dot line), the equality of the cycles prevents a good match to the sides and other regions of cyclic variation. The resulting turbulence fluctuation signal again has inconsistent cyclic variation, with relatively high values during the steep jet velocity portions, and significant fluctuations during the low jet velocity values (turbulence intensity of 0.839 m/s).

b. Case 2: high jitter (J=5.9%). Figure 6 constitutes a case with fairly high jitter of 5.9%. Smoothing (solid line) shows a reasonable deterministic fit, with turbulence intensity of 0.870 m/s. The ensemble averaging shows a poor fit, large turbulence fluctuation values, and a high turbulence intensity of 3.660 m/s. Figure 6 indicates that the wavelet denoising was comparable to the smoothing technique. In all methods, the turbulent fluctuations have their maximum values at the instants of opening and closing. This is probably due to the amplification of the noise by the large velocity gradient at the instants of opening and closing that creates some technical difficulty in resolving glottal turbulence and may reduce the reliability of every method at those instants. These peaks also grow larger as the jitter value increases.

c. Case 3: very high jitter (J=12.4%). This case is shown in Fig. 7. The smoothing (solid line) shows the best deterministic fit and lowest turbulence intensity (0.889). The wavelet denoising does not appear to perform as well as the smoothing (intensity=1.581), and the turbulence fluctuations have higher values. The ensemble average (dash-dot-dot) does not fit well and has extremely high turbulence fluctuations (intensity=5.001). It is noted that the judgment of how well the deterministic signal appears to fit the original signal is performed visually, noting especially such aspects as a smooth following of the gross changes of the data points. It is also noted that the original (and deterministic) velocity signals have double peaks, which may be due to either vocal fold motion anomalies and/or movement of the glottal jet center back and forth away from or across the centerline once within each cycle, or possibly from subglottal resonance effects on the transglottal pressure that drives the air. Double peaks in the velocity signal have been noted before (Berke *et al.*, 1989; Alipour and Scherer, 1995).

Table I shows that the turbulence intensity increases with jitter for the ensemble-averaging method, due to individual cycle shapes becoming more different from the single ensemble-averaged cycle as jitter increases. The turbulence intensity does not increase with jitter for the other two methods. It is noted that jitter itself is based on the EGG signal, not on the microphone signal, and thus refers to vocal fold motion aspects. Therefore, the turbulence intensity appears to reflect airflow phenomena rather than tissue kinematics irregularities, an important distinction relative to acoustic implications. This is made clearer by the understanding that glottal air velocities are much faster than vocal fold tissue

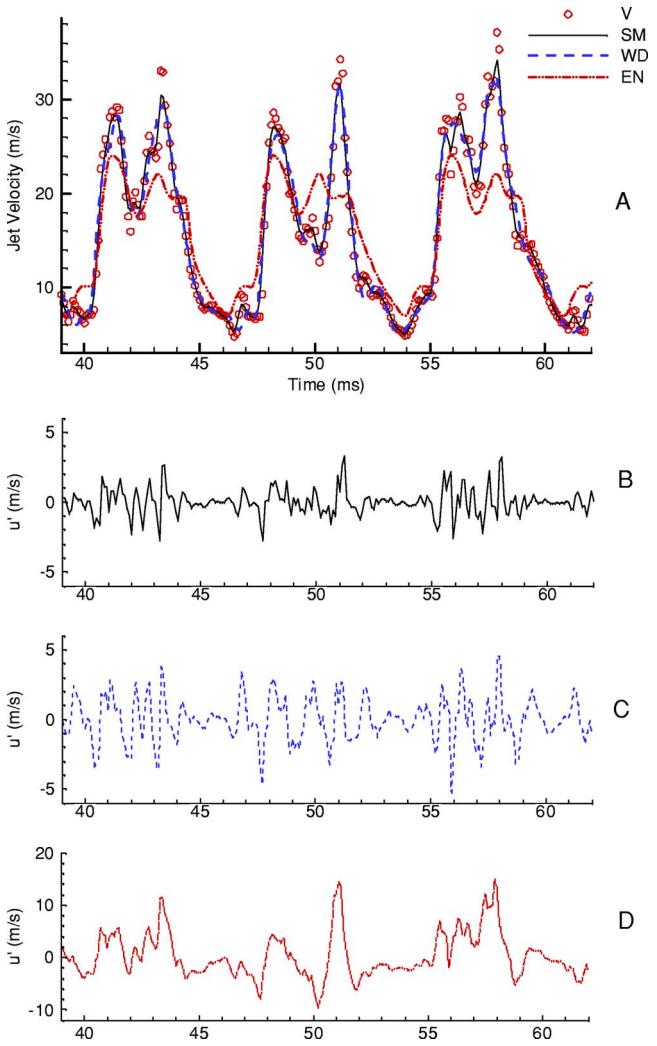


FIG. 7. (Color online) Turbulence averaging for a very high-jitter larynx (12.35%). The descriptions of the signals are the same as Fig. 3.

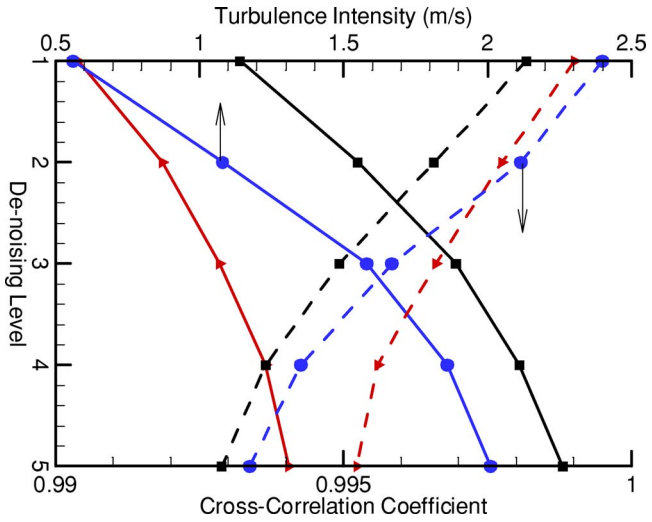


FIG. 8. (Color online) Effects of the wavelet denoising level on the turbulence averaging. The solid lines represent turbulence intensity and dashed lines represent cross-correlation coefficient of the denoised and original signal for the three larynges of Table I.

velocities, and jet vortex structures move away from the glottis at much higher rates than the tissue motion (Kucinschi, 2004).

In order to examine the effect of choosing different "levels" for the wavelet denoising technique, five cases are shown in Fig. 8. In this figure, the turbulence intensity (solid lines) and the cross-correlation between the deterministic and original velocity signals (dashed lines) are graphed against the denoising level. It is seen that by increasing the denoising level, the turbulence intensity *increases*, suggesting (1) higher denoising levels create poorer deterministic fits to the original signal, and (2) the choice of which level to use to describe turbulence intensity for phonatory signals appears arbitrary, and therefore difficult if not impossible to set. Figure 8 also indicates that the cross-correlations decrease as denoising increases.

2. Spectral comparisons

Another important aspect of the process of separating the deterministic signal from the turbulence is the power spectrum of the deterministic and fluctuation velocities from each method. Figures 9(a)–9(c) show these spectra for the three methods for a low-jitter (0.7%) dataset, collected on larynx #6 with a sampling rate of 20 000 samples/s. The spectra were taken over a period of 4 s of phonation at 208 Hz. The spectra were computed in MATLAB with the Welch method, a Hamming window, a FFT size of 4096 pints, and a segment size of 64 with 50% overlap. The power spectra of all three methods for the deterministic and the fluctuating signals showed obvious cyclic structure. The spectra from the smoothing method [Fig. 9(a)] shows an overlap around 3500 Hz and spikes disappear after that frequency. The wavelet method [Fig. 9(b)] shows a similar behavior, with a sharp increase of energy of turbulence until the crossing point at around 1400 Hz. However, the deterministic signal shows stronger spikes at higher frequencies, suggesting that, in this context, the wavelet method may overfit the signal in the high frequencies, creating a deterministic signal into higher frequencies that may be unrealistic. The ensemble-average spectra [Fig. 9(c)] are quite different from the other two in that there are sharp spikes of the turbulent fluctuations at all frequencies, and they are very strong at low frequencies. The existence of the same frequency structure in the turbulent fluctuation velocity indicates that glottal turbulence has intermittency. That is, it is prominent when the glottis opens and diminishes as it closes.

3. Turbulence intensity related to subglottal pressure

It would be reasonable to assume that turbulence intensity would increase with subglottal pressure and with less prephonatory glottal adduction, because in both cases greater flow passes through the glottis, at least at certain times of the phonatory cycle. The data of Fig. 10 tend to support these hypotheses. For the larynx shown in Fig. 10, three levels of adduction were used (shims of different sizes were placed between the arytenoid cartilages to establish prephonatory adduction levels). On average (from linear regression), the turbulence intensity increased approximately 0.055 m/s per cm H_2O (vocal fold length was held constant). The results suggest that for a given adduction level, increased subglottal

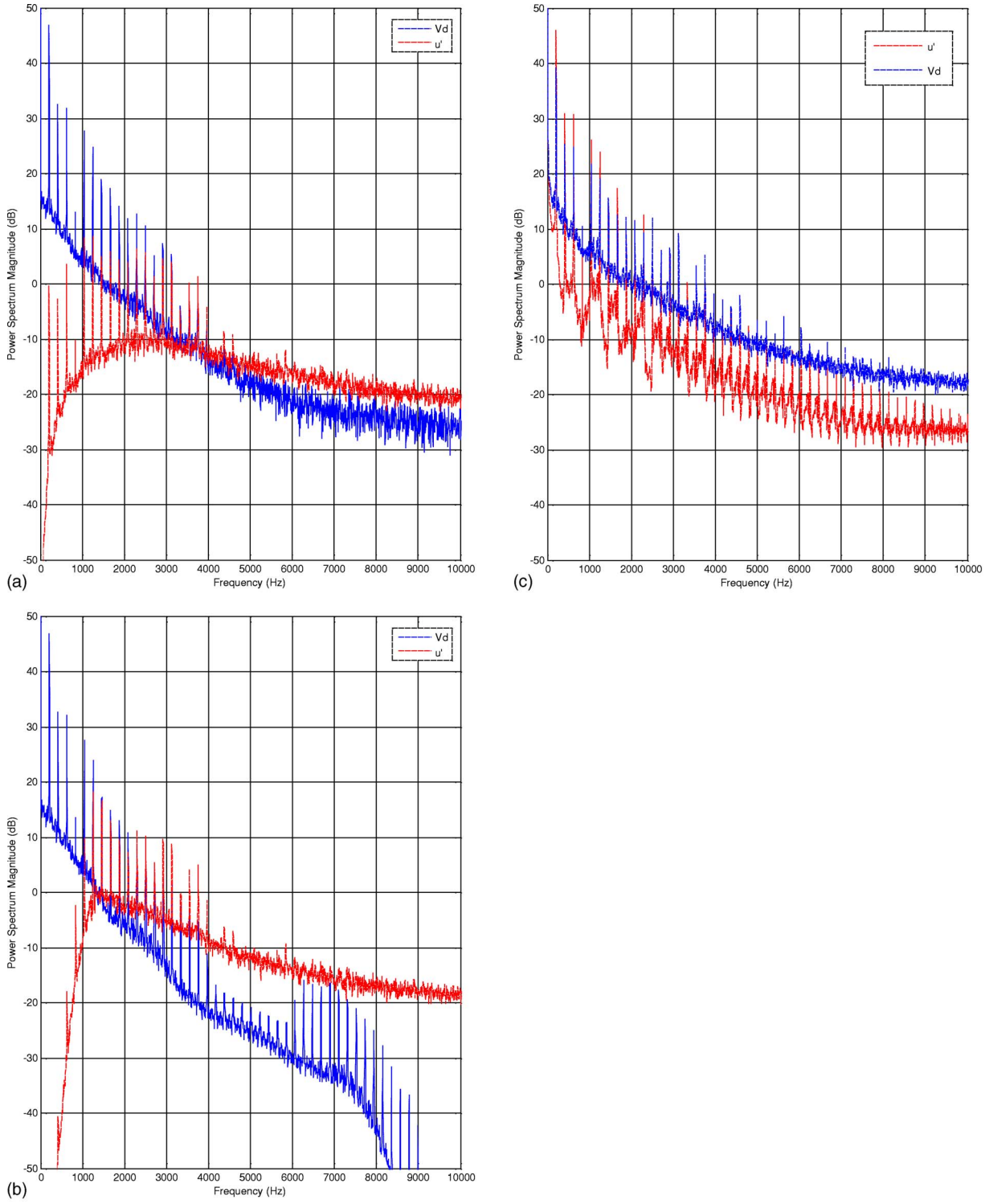


FIG. 9. (Color online) Power spectra comparison of the deterministic and turbulent fluctuation velocities for the three methods. (a) smoothing method; (b) for wavelet denoising, and (c) for the ensemble-averaging method.

pressure tends to increase the turbulence intensity. As the adduction decreased, turbulence intensity tended to increase. This was caused by the increase of flow rate due to lower flow resistance.

4. Turbulence intensity versus Reynolds and Strouhal numbers

Figures 11 and 12 show the variations of the turbulence intensity as a function of Reynolds and Strouhal numbers, which are the major aerodynamic parameters for the oscillating larynx. The turbulence intensities for these graphs were calculated from the smoothing filter method. These graphs show the turbulence intensities for four larynges (each symbol represents one larynx). While each larynx created a cluster of data, for each larynx there was an increasing trend of turbulence intensity both with Reynolds number and with Strouhal number. The range of turbulence intensity was from about 0.5 to 3.5 m/s for these canine larynges. These data

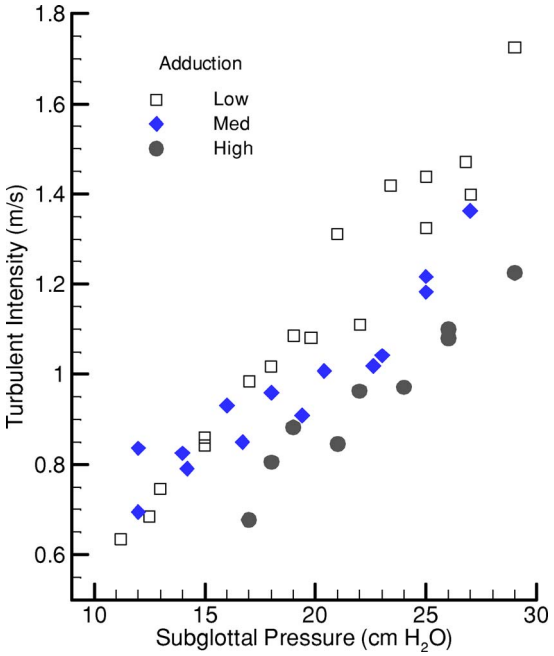


FIG. 10. (Color online) Turbulent intensities in an oscillating larynx with various adduction levels as a function of subglottal pressure.

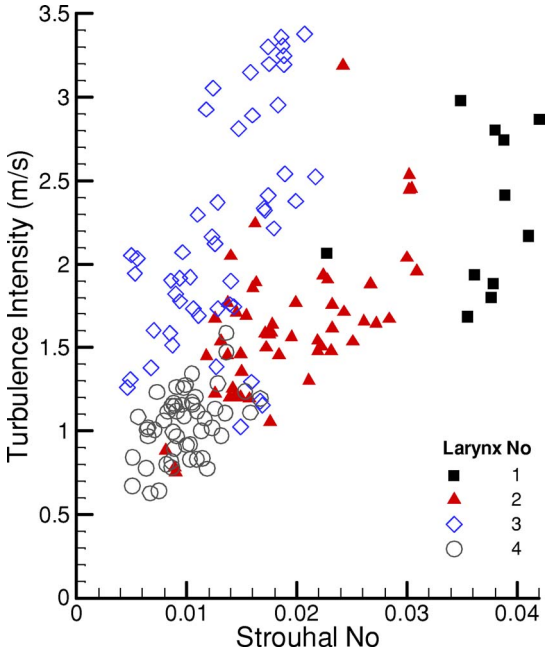


FIG. 12. (Color online) Turbulent intensity as a function of Strouhal number for four excised larynges.

points represent a wide range of oscillations of the canine larynges at various adduction and pressure flow values. Thus large variability in the data may be attributed to the difference in the larynx and adduction conditions.

IV. SUMMARY

The existence of flow turbulence in the glottal jet flow has been shown to be highly likely. The importance of the turbulence has not been well established, although it may appear obvious that it is part of the noise creation for breathy voicing, whisper, and perhaps secondary dipole sound. Thus, it is important to establish methods to measure turbulence associated with phonation. Turbulence intensity is the conventional measure of choice. But what methods should be used to extract the turbulence from velocity signals? In this study we attempted to examine three different methods, smoothing, wavelet denoising, and ensemble averaging.

The results suggest that the ensemble averaging is inadequate because it gives the worst representation of the deterministic signal and the lowest cross-correlation between the deterministic signal and the original velocity. It takes away cycle "individuality" by making them all the same (for the chosen group with which to calculate the deterministic signal). The wavelet denoising scheme would appear to be highly applicable, but the requirement to make a choice of which level of analysis to make lends flexibility for which there is an absence of knowledge. That is, the actual turbulence intensity is unknown, and therefore the level of analysis is not determinable. The smoothing method appears to be a reasonable compromise at this time because of its reasonable calculation of both the deterministic signal and the turbulence intensity, and because of its simplicity.

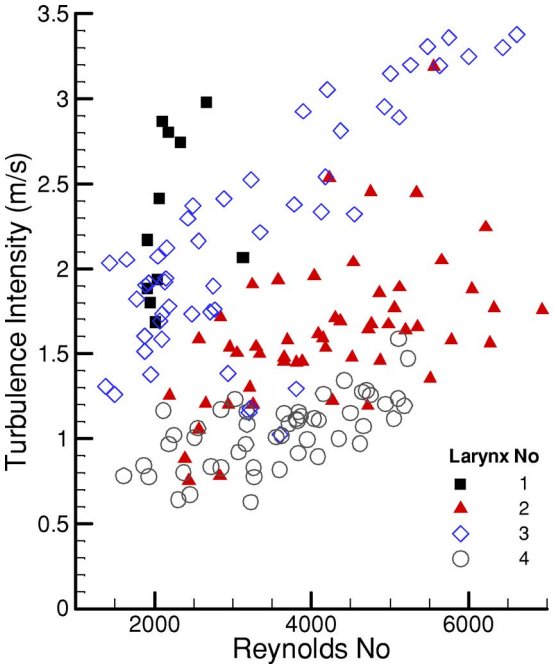


FIG. 11. (Color online) Turbulent intensity as a function of Reynolds number for four excised larynges.

V. CONCLUSIONS

Three methods of extracting both the deterministic velocity signal and the fluctuating velocity signal from phonating excised larynxes were compared. The *ensemble-averaging* method would work well if the deterministic signal were periodic, but in phonation this is rarely the case, and thus the deterministic signal lacks the expected gross cyclic perturbations and the calculated turbulence intensity is exaggerated. The *wavelet denoising* method can perform an excellent analysis and synthesis, but for determining an accurate representation of turbulence and turbulence intensity, there yet are no guidelines for determining which analysis

levels to use. The *smoothing* method used, whereby a sequential weighted smoothing takes place, preserves important gross features of the deterministic signal and provides a reasonable measure of turbulence intensity. Because it is imperative to preserve the gross features of the phonatory deterministic signal, this study suggests that the smoothing technique should be considered when extracting turbulence intensity from phonatory velocities. The simplicity of this method is an added advantage. A further evaluation of similar methods for phonation may prove useful. It is stressed that at this time any method that appears to be best for phonation purposes will be related to the issues of sound production, vocal fold oscillation, and intervention strategies for phonatory improvement, and thus comparative studies of various methods should continue. Although the measured data were for the open jet and the actual glottal jet must pass between false vocal folds creating a realistic complication in modeling that needs to be addressed, it is expected that the application of extraction of turbulence measures can aid both aerodynamic research with laryngeal models as well as specific clinical problems with humans. It is conceivable that velocity measures at the laryngeal level will be feasible in the future when accessibility and size requirements are met.

ACKNOWLEDGMENTS

This work was supported by Research Grant R01 DC03566 from the National Institute on Deafness and other Communication Disorders. The authors would like to thank Dr. Anders Löfqvist and an anonymous reviewer for their helpful comments.

Alipour, F., and Scherer, R. C. (**1995**). "Pulsatile airflow during phonation: An excised larynx model," J. Acoust. Soc. Am. **97**, 1241–1248.

Alipour, F., Scherer, R. C., and Patel, V. C. (**1995**). "An experimental study of pulsatile flow in canine larynges," J. Fluids Eng.. **117**, 577–581.

Alku, P., and Vilkman, E. (**1996**). "A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers," Folia Phoniatr Logop **48**, 240–254.

Baldwin, J. T., Deutsch, S., Petrie, H. L., and Tarbell, J. M. (**1993**). "Determination of principal Reynolds stresses in pulsatile flows after elliptical filtering of discrete velocity measurements," J. Biomech. Eng. **115**, 396–403.

Berke, G. S., Moore, D. M., Monkewitz, P. A., Hanson, D. G., and Gerratt, B. R. (**1989**). "A preliminary study of particle velocity during phonation in an in-vivo canine model," J. Voice **3**, 306–313.

Bluestein, D., and Einav, S. (**1995**). "Effect of varying degrees of stenosis on the characteristics of turbulent pulsatile flow through heart valves," J. Biomech.. **28**, 915–924.

Childers, D. G., and Ahn, C. (**1995**). "Modeling the glottal volume-velocity waveform for three voice types," J. Acoust. Soc. Am. **97**, 505–519.

Childers, D. G., and Lee, C. K. (**1991**). "Vocal quality factors: analysis, synthesis, and perception," J. Acoust. Soc. Am. **90**, 2394–2410.

Cooke, A., Ludlow, C. L., Hallett, N., and Selbie, W. S. (**1997**). "Characteristics of vocal fold adduction related to voice onset," J. Voice **11**, 12–22.

de Krom, G. (**1995**). "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," J. Speech Hear. Res. **38**, 794–811.

Einav, S., and Sokolov, M. (**1993**). "An experimental study of pulsatile pipe flow in the transition range," J. Biomech. Eng. **115**, 404–411.

Hamming, R. W. (**1973**). *Numerical Methods for Scientists and Engineers* (McGraw-Hill, New York).

Hillenbrand, J., and Houde, R. A. (**1996**). "Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech," J. Speech Hear. Res. **39**, 311–21.

Hillman, R. E., Oesterle, E., and Feth, L. L. (**1983**). "Characteristics of the glottal turbulent noise source," J. Acoust. Soc. Am. **74**, 691–694.

Hofmans, G. C. J., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). "Unsteady flow through in-vitro models of the glottis," J. Acoust. Soc. Am. **113**, 1658–1675.

Hussain, A. K. M. F (**1986**). "Coherent structures and turbulence," J. Fluid Mech. **173**, 303–356.

Kucinschi, B. R. (2004). "An analysis of the flow through a driven mechanical model of the vocal folds," Ph.D. thesis, The University of Toledo, Toledo, Ohio.

Kucinschi, B. R., Scherer, R. C., De Witt, K. J., and Ng, T. T. M. (**in review**). "Flow visualization of the air moving through a model of the larynx," J. Fluids Eng.

Resnikoff, H. L., and Wells, R. O. (**1998**). *Wavelet Analysis—The Scalable Structure of Information* (Springer-Verlag, New York).

Shinwari, D., Scherer, R. C., De Witt, K. J., and Afjeh, A. A. (**2003**). "Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees," J. Acoust. Soc. Am. **113**, 487–497.

Shrivastav, R. and Sapienza, C. M. (**2003**). "Objective measures of breathy voice quality obtained using an auditory model," J. Acoust. Soc. Am. **114**, 2217–2224.

Sullivan, P., Ancimer, R., and Wallace, J. (**1999**). "Turbulence averaging within spark ignition engines," Exp. Fluids **27**, 92–101.

Titze I. R., Horii, Y., and Scherer, R. C. (**1987**). "Some technical considerations in voice perturbation measurements," J. Speech Hear. Res. **30**, 252–260.

Walburn, F. J., Sabbah, H. N., and Stein, P. D. (**1983**). "An experimental evaluation of the use of an ensemble average for the calculation of turbulence in pulsatile flow," Ann. Biomed. Eng. **11**, 385–399.

White, F. M. (**1974**). *Viscous Fluid Flow* (McGraw-Hill, New York).

Wolfe, V., and Martin, D. (**1997**). "Acoustic correlates of dysphonia: type and severity," J. Commun. Disord. **30**, 403–415.

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.

The effect of intertalker speech rate variation on acoustic vowel space

Ying-Chiao Tsao[a)]
Human Communication Studies/Communicative Disorders, 800 North State College Boulevard, P.O. Box 6868, California State University-Fullerton, Fullerton, California 92834-6868

Gary Weismer
Waisman Center and Department of Communicative Disorders, 1500 Highland Avenue, University of Wisconsin-Madison, Madison, Wisconsin 53705-2280

Kamran Iqbal
Department of Systems Engineering, University of Arkansas at Little Rock, 2801 South University Avenue, Little Rock, Arkansas 72204-1099

(Received 27 June 2005; accepted 11 November 2005)

The present study aimed to examine the size of the acoustic vowel space in talkers who had previously been identified as having slow and fast habitual speaking rates [Tsao, Y.-C. and Weismer, G. (1997) J. Speech Lang. Hear. Res. **40**, 858–866]. Within talkers, it is fairly well known that faster speaking rates result in a compression of the vowel space relative to that measured for slower rates, so the current study was completed to determine if the same differences in the size of the vowel space occur across talkers who differ significantly in their habitual speaking rates. Results indicated that there was no difference in the average size of the vowel space for slow vs fast talkers, and no relationship across talkers between vowel duration and formant frequencies. One difference between the slow and fast talkers was in intertalker variability of the vowel spaces, which was clearly greater for the slow talkers, for both speaker sexes. Results are discussed relative to theories of speech production and vowel normalization in speech perception. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149774]

PACS number(s): 43.70.−h, 43.70.Fq, 43.70.Gr, 43.70.Bk [AL] Pages: 1074–1082

I. INTRODUCTION

The acoustical realization of speech sounds is known to be associated with a great deal of variability, even in the presence of perceptual constancy (Perkell and Klatt, 1986). One of the aims of research on speech production and its application to speech technology is to place this variability into an explanatory organizational framework (Lindblom, 1990). One element of this explanatory framework involves the delineation of the influence of suprasegmental variables on the spectral attributes of speech sounds. A concerted effort over the last half of the previous century has focused on the study of the relationship of vowel formant frequencies to speaking rate variation (Miller, 1981; Miller, O'Rourke, and Volaitis, 1997).

Stetson (1951) was one of the first investigators to note a possible relationship between temporal and spectral aspects of vowel production. Based on perceptual observations, Stetson claimed that vowels in unstressed syllables became schwalike as speaking rate increased. Later, Tiffany (1959) noted changes in vowel formant frequencies with changes in level of stress. The vowel space, as enclosed by the planar coordinates of the first and second formant frequencies of selected vowels, was reduced in size when vowels were produced in an unstressed, as compared to a stressed syllable. This observation is relevant because decreases in stress are typically associated with decreases in vowel duration (Fry, 1955). Thus, the results of both Stetson and Tiffany point to a possible role of vowel duration in the spectral—and possibly perceptual—characteristics of vowels.

Stevens and House (1963) investigated the influence of phonetic environment on formant frequencies of American vowels, and found that of their three speakers the one with the shortest vowel duration typically had the greatest F2 difference between a vowel in "null" context (/hVd/) and the same vowel in the consonant context. These authors speculated that greater spectral undershoot may occur for speakers exhibiting a faster speaking rate due to the adoption of speaker-specific movement characteristics or strategies. Apparently, no further work has been reported on a possible link between rate differences across speakers and their associated vowel spaces.

There have been continued efforts toward understanding the influence of speaking rate—within speakers—on the spectral and temporal properties of vowel production (Fourakis, 1991; Gay, 1978; Klatt, 1976; Lehiste, 1970; Stevens and House, 1963). Lindblom (1963), employing a procedure designed to elicit a large range of vowel durations in CVC syllables produced by a single speaker of Swedish, found a slight change in the first and third formant frequencies as a function of vowel duration, but a relatively large effect for the second formant. As vowel duration decreased, formant frequencies tended toward the F-pattern for schwa and away from the so-called, idealized targets. Similar find-

[a)]Electronic mail: ytsao@fullerton.edu

ings of vowel reduction with increases in speaking rate have been reported for Japanese vowels by Tsukada and Hirata (2003).

Additional studies on a possible speaking rate effect on the size of the acoustic vowel space produced a somewhat mixed picture. In a study examining the effects of changes in stress and rate of speech (tempo) on acoustic characteristics of American English vowels produced by four males and four females in two contexts (i.e., [hVd] and [bVd]), Fourakis (1991) reported that the size of the vowel space was reduced by 30% from the slow-stressed to the fast-unstressed condition. This result is in agreement with Miller (1981) who reported that fast tempo and less stress bring about shrinkage of the overall vowel space. van Son and Pols (1990) studied seven Dutch vowels produced in a meaningful text of about 850 words read twice, once at normal rate and once at fast rate, by an experienced newscaster. They did not observe changes in F1/F2 values as a function of rate, as found in the previous studies (Gay, 1978, Gopal and Syrdal, 1988; Engstrand, 1988). van Son and Pols (1990) indicated that the only systematic effect was a higher F1 in fast-rate speech, for all vowels, that may suggest a generally more open articulation of vowels as a result of speaking louder or some other general change in speaking style induced by the requirement to speak rapidly. These results are not easily fit in to expectations for a relationship between rate and formant frequencies because the data are limited to a single, professional speaker. What is clear from van Son and Pols' (1990) work is that changes in vowel duration induced by modification of speaking rate do not necessarily have to produce changes in vowel formant frequencies. The strength and frequency of relationships between vowel duration and vowel formant frequencies therefore remains a topic for debate.

Lindblom (1963), Moon and Lindblom (1994), and Turner, Tjaden, and Weismer (1995) have reported relatively strong intratalker relationships between rate and the size of the vowel space, whereas Fourakis (1991), Gay (1978), van Bergem (1993), and van Son and Pols (1990) have reported lesser or absent effects. The weight of the evidence, however, seems to be that the acoustic vowel space tends to be smaller at fast, as compared to slow rates (Turner, Tjaden, and Weismer, 1995; see also, Weismer, Laures, Jeng, Kent, and Kent, 2000, Figures 3 and 4, pp. 210–211), the latter two conditions being associated with systematically different vowel durations.

Variation in speech rate is known to exist within and among speakers (Crystal and House, 1982; Miller, 1981; Miller, Grosjean, and Lomanto, 1984; Tsao and Weismer, 1997; Turner and Weismer, 1993). More specifically, several factors are known to contribute to systematic variation in the temporal characteristics of speech sounds, including: *the speech context* (in isolation, syllable structures, connected speech, utterance length, e.g., Fourakis, 1991; Hirata and Tsukada, 2004; Klatt, 1976; Lindblom, 1963; Moon and Lindblom, 1994; Tuller, Harris, and Kelso, 1982; Tsukada and Hirata, 2003), *the speaker* (Kuehn and Moll, 1976; Ostry and Munhall, 1985; Johnson, Ladefoged, and Lindau, 1993; Xue and Hao, 2003), *the rate of speech* (Miller, 1981; Miller and Volaitis, 1989; Miller, Grosjean, and Lomanto, 1984) and others. Smith (2002) reported considerable variability among 15 participants in his normal speaking rate condition in the extent to which temporal parameters such as final-syllable vowel lengthening occurred. In addition, there was substantial variation in these parameters in the fast speaking rate condition, and some systematic changes in certain patterns also occurred as a function of the rate change.

In an attempt to explore the underlying mechanisms for variations in an individual's speech rate, Tsao and Weismer (1997) studied the habitual and maximal speaking rates of one hundred talkers reading the Farm Passage (Crystal and House, 1982). Passage duration at a self-selected habitual rate was measured, and each participant was classified into one of three rate groups: average, slow, and fast, based on the measured durations. More specifically, the SLOW group had average syllable durations that were 73 and 48 ms longer than the syllable durations of the FAST group when expressed as speaking rate and articulation rate, respectively (see Tsao and Weismer, 1997, Tables 1 and 2, p. 862). Most interestingly, the average *habitual* articulation rate of the fast speakers was about the same as the average *maximum* articulation rate of the slow speakers. Tsao and Weismer (1997) therefore suggested that interspeaker variation in habitual speaking rate is predictive of interspeaker variation in maximal speaking rate. Speakers who are habitually slow do not seem to be able to produce as fast a maximum rate as speakers who are habitually fast. This finding was interpreted to mean that interspeaker variation in habitual speaking rate might reflect neurological predispositions rather than (for example) voluntary "choices" of different rates. If this speculation is correct, other differences in the speech production of slow vs fast talkers might also be hypothesized. Based on previous literature, perhaps the vowel spaces of habitually slow and fast talkers are different in ways predictable from intratalker rate effects.

As noted above, to date all experimental rate manipulations for which the effect of relatively compressed vowel spaces at faster as compared to slower speaking rates have been of the *within*-talker variety. Tsao and Weismer's (1997) demonstration of an across-speaker, categorical difference between fast and slow talkers prompts the following question: If, in fact, there is a real difference between habitually slow and fast talkers, do these two groups of speakers produce different-sized vowel spaces, and specifically do habitually fast talkers produced relatively compressed vowel spaces relative to habitually slow talkers? This question is not only important for understanding talker differences and their role in theories of speech production, but also may bear on the use of the vowel space as an index of speech motor involvement in various speech disorders.

II. METHODS

A. Participants

The study included 30 participants (15 slow and 15 fast talkers), ranging in age from 18 to 35 years, selected from 100 participants (50 males, 50 females) from a previous study (Tsao and Weismer, 1997). All speakers had dialects heard in the upper Midwest of the USA, and most grew up in

Wisconsin and Illinois. In Tsao and Weismer (1997) participants were asked to read the Farm Script (Crystal and House, 1982) at their habitual and maximal speaking rates. Based on total reading time and overall speaking rate of each participant's performance on this passage, talkers were classified into slow and fast groups. Essentially, the slow talkers were chosen from one end of the speaking rate distribution for the 100 original participants, and the fast talkers from the other end (Tsao and Weismer, 1997). Among the 15 speakers per rate group, there were 7 males and 8 females in the Slow group and 8 males and 7 females in the fast group. Data reported below show clearly that the vowel durations of talkers in these two rate groups were systematically different.

B. Data collection

The Farm Script reading (Crystal and House, 1982) of each participant was recorded using a high quality audiotape recorder (Tandberg, TCD 420A) equipped with a Shure 1800 head-mounted microphone. Seven complete readings of the Farm Script were recorded for each participant at a participant-selected, habitual speaking rate. Vowels used for acoustic measures included the corner vowels /i/, /æ/, /u/, and /a/ in monosyllabic words of CVC, CV, VC, VCC, CVCC, and CCVCC form. Multiple productions of each vowel (i.e., 98 /i/, 266 /æ/, 119 /u/, and 70 /a/) produced by each participant were identified among the 327 vowels in the Farm script (Crystal and House, 1982). A few of the vowels used for the analysis came from function words such as "too" (for /u/) and "be" (for /i/), but the majority were from content words. Some vowels were excluded from the analysis because they were preceded or followed by /r/, /l/, or /w/, which created segmentation difficulties. Ultimately, a total of 36 CVC, 17 CV, 10 VCC, 5 VC, and 11 CVCC and CCVCC tokens were analyzed. Within these syllable shapes, 79 vowels (10 /a/, 17 /u/, 38 /æ/, and 14 /i/) from each repetition of the Farm script passage were measured, yielding a total of 553 vowels per talker for 7 repetitions.

C. Acoustic measures and analysis

Speech samples were digitized and measured at a sampling rate of 22 kHz and low-pass filtered at 9.8 kHz using CSpeech (Milenkovic, 1994). Conventional segmentation criteria (Weismer, 1984) were used to measure vowel segment durations, that is, the time interval between the first and last glottal pulse of a vocalic nucleus was considered to be the vowel duration. Vowel formant frequency measures were made using both wideband spectrogram (300-Hz filter for males, 450 Hz filter for females) and linear predictive coding (LPC, 24 coefficients) displays generated by CSpeech. F1 and F2 frequencies were measured, within a 30-ms Hamming window centered at the temporal midpoint of the vowel, by identifying the peaks in the LPC spectrum and checking those against the digital spectrogram. When an LPC peak was clearly at odds with the cursor location in the spectrographic display, the adjustment was made using the spectrogram as the standard (i.e., the frequency cursor was placed in the center of the formant band bounded by the 30 ms window). For vowels shorter than 30 ms, the analysis window was considered to be the entire vowel.

Based on these measurements, an average value for F1 and F2 was computed for each vowel produced by each individual talker. An acoustic vowel quadrilateral, derived from the corner vowels, was then constructed for each talker. The area of the acoustic vowel space (expressed in Hz^2) enclosed by this quadrilateral was automatically calculated and plotted using the polygon area function in Matlab software (Mathworks, Inc., 2000), with the average formant frequencies serving as input. Computed vowel space areas were separated by rate group (slow vs fast) and, within these groups, by gender.

TABLE I. Means and standard deviations of vowel duration (in ms) for /i/, /æ/, /u/, and /a/ for talkers within the rate groups (slow vs fast) originally identified by Tsao and Weismer (1997). t-test results are shown below each vowel.

	/i/	/æ/	/u/	/a/
Slow (7M+8F)	129 (16)	154 (17)	110 (16)	138 (22)
Fast (8M+7F)	97 (16)	124 (19)	86 (11)	112 (19)
	$t(28)=5.477$[a]	$t(27)=4.557$[a]	$t(24)=4.787$[a]	$t(27)=3.464$[a]

[a]All one-tailed tests significant at $p<0.001$, with degrees of freedom adjusted within each comparison for group variances.

D. Reliability of acoustic measures

To obtain intrajudge reliability, four corner vowels of eight participants (4 slow, 4 fast; 5 males and 3 females) were randomly selected from the collected speech samples, resulting in 10% (55 vowels of 553 per talker) of the total data set. The vowel durations and formant frequencies were re-examined using the same procedures described above. The correlations between the first and second set of segment duration measures were 0.95 for vowel durations, 0.99 for F1, and 0.98 for F2.

III. RESULTS

A. Vowel durations

Table I reports mean vowel durations for the two rate groups. The average durations of all four vowels were longer for the slow, as compared to the fast group, with group differences ranging from 24 ms for /u/ to 32 ms for /i/. As indicated in Table I, each vowel difference across the groups was statistically significant when evaluated by one-tailed t-tests. If there is an across-speaker effect of habitual speaking rate on the acoustic vowel space, these vowel duration data demonstrate that the effect should be revealed by this sample.

B. Formant frequencies and acoustical vowel spaces

Table II reports mean F1 and F2 values for the current talkers, separated by gender and pooled across rate groups, and also includes data for comparison from Hillenbrand, Getty, Clark, and Wheeler (1995); Table III reports current formant frequency data separated by the two rate groups. The formant frequencies measured in the present investigation

TABLE II. Means and standard deviations of formant frequencies (in Hz) for vowels /i/, /æ/, /u/, and /a/ for talkers pooled across rate groups but separated by gender; comparison mean data from Hillenbrand *et al.* (1995) are shown on the right-hand side of the table.

	Current study				Hillenbrand *et al.* (1995)			
	F1		F2		F1		F2	
	Male	Female	Male	Female	Male	Female	Male	Female
/i/	377	453	2015	2371	342	437	2322	2761
	(17)	(37)	(106)	(123)				
/æ/	628	793	1706	2021	588	669	1952	2349
	(63)	(61)	(69)	(130)				
/u/	391	474	1444	1675	378	459	997	1105
	(25)	(43)	(114)	(146)				
/a/	713	880	1431	1674	768	936	1333	1551
	(47)	(57)	(58)	(105)				

compare favorably, in most cases, to those reported by Hillenbrand *et al.* (1995) for the so-called "null" context. Many of the larger differences between the two data sets (such as the F2's for /i/, /u/, and /æ/) are easily explained on the basis of the tendency for formant frequencies to move from their 'null' context values toward more centralized values when produced in "full" context (e.g., Stevens and House, 1963; Hillenbrand, Clark, and Nearey, 2001), which describes the speech sample used in the present investigation. Table III shows the essentially identical average F-patterns of speakers in the slow and fast groups. Despite the clear differences in vowel duration between these groups (Table I), their average formant frequencies for the corner vowels appear to be largely the same.

Although statistically identical vowel space areas do not follow unequivocally from the finding of nearly equivalent formant frequencies for the two rate groups (small differences between individual formants could accumulate in a consistent direction to produce a systematic area difference), data reported in Table IV show that the size of the acoustic vowel space was basically the same for the slow and fast talkers; within-gender t-tests [$t(13)=0.266$, $p=0.7949$ for males; $t(13)=0.993$, $p=0.3375$ for females] proved the statistical equivalence. The larger vowel space areas for females, as compared to males, are most likely due strictly to the different frequency ranges occupied by the formant frequency data rather than a reflection of gender-based articulation styles (see also, Weismer, Kent, Hodge, and Martin, 1988; and related data in Simpson, 2001).

One difference between the vowel space areas of slow and fast talkers is that intertalker variability is greater in the slow group, for both males and females. This is seen in Fig. 1, where individual data points in F1–F2 plots are shown by rate group and gender for each vowel; the solid line connects the average coordinate points for the four corner vowels, and therefore defines the average vowel space. The intertalker variability issue is taken up in the discussion.

C. The relationship between vowel durations and formant frequencies

A series of regression analyses was conducted to explore potential relationships between vowel durations and formant frequencies. Although there were no differences between the formant frequencies and vowel space areas of slow and fast talkers, relationships might still be shown for selected groups or vowels. Toward that end, regression analyses were conducted within the two groups (slow and fast) with data pooled across vowels, as well as within the two groups for each of the four vowels. In both the pooled-vowel and individual-vowel analyses, the data pairs were the mean vowel durations and mean formant frequencies for an individual speaker. Thus the pooled-vowel regressions were based on 60 points (15 speakers per group $\times 4$ vowels) and the individual-vowel analyses on 15 points.

Of the four pooled-vowel analyses (slow-F1, slow-F2, fast-F1, fast-F2), the regression functions relating vowel duration to F1 were significant for both the slow [$F(2,58)=16.72$, $p=0.0001$, $r^2=0.22$] and fast [$F(2,58)=21.83$, $p<0.0001$, $r^2=0.27$] groups; neither function for vowel duration and F2 was significant. In addition, none of the

TABLE III. Means and standard deviation of formant frequencies (in Hz) for vowels /i/, /æ/, /u/, and /a/ for talkers separated by rate group (slow vs fast).

	/i/		/æ/		/u/		/a/	
	F1	F2	F1	F2	F1	F2	F1	F2
Slow (7M, 8F)	406	2200	702	1862	430	1552	797	1541
	(30)	(122)	(71)	(91)	(36)	(156)	(76)	(85)
Fast (8M, 7F)	424	2186	719	1865	435	1567	796	1564
	(24)	(107)	(54)	(108)	(32)	(104)	(29)	(78)

TABLE IV. Means and standard deviation of the size of vowel space areas (in Hz^2) encompassed by vowels (/i/, /æ/, /u/, and /a/), reported by talker group (slow vs fast) and gender.

	Slow	Fast	Mean (SD)
Male	114136 Hz^2 (SD=31120)	111837 Hz^2 (SD=12333)	112910 (22193)
Female	197303 Hz^2 (SD=72872)	175682 Hz^2 (SD=42513)	187213 (61765)
Mean (SD)	158492 (70104)	141631 (44018)	

individual-vowel analyses had significant regression functions between vowel duration and either F1 or F2.

Plots of the two significant pooled functions are shown in Fig. 2, in which the four vowels are coded by separate symbols. Filled points plot data for high vowels, unfilled points for low vowels. These plots demonstrate that the significant regression functions are artifacts, in a sense, of the covariance for high and low vowels between vowel duration and the height of the first formant. High vowels tend to have shorter durations than low vowels, and lower F1's as well; in both plots the filled symbols tend to occupy the left, lower quadrant, the open symbols the right, upper quadrant. The significant functions do not appear to reveal any deterministic relationship between the average duration of a talker's vowels and his or her average F1's.

IV. DISCUSSION AND CONCLUSIONS

The present study was motivated by consideration of two empirical phenomena. First, there have been several studies in the literature (Lindblom, 1963; Miller, 1981; Turner, Tjaden, and Weismer, 1995; Weismer, Laures, Jeng, Kent, and Kent, 2000) showing, within speakers, a compression of the acoustic vowel space with increased speaking rate. Second, Tsao and Weismer (1997) expanded on the well-known *across*-speaker variation in habitual speaking rate (e.g., Miller, Grosjean, and Lomanto, 1984) by demonstrating that, among a relatively large group of talkers ($N=100$), a "slow" group could be distinguished statistically from a "fast" group and that talkers within these respective groups had statistically different *maximum* speaking rates. The slower maximum rates of the slow, as compared to fast, talkers suggested differences that were more likely driven by neurological variation than speaker-specific "selection" of habitual speaking rates. The current study coordinated these two findings by investigating possible differences in the size of the acoustic vowel space for speakers in the slow and fast groups. A reasonable hypothesis seemed to be that fast talkers would have smaller vowel spaces as compared to slow talkers. Moreover, the logic of, and motivation for, the hypothesis can be extended to the issue of rate normalization as a front-end process in speech perception (e.g., see Miller and Dexter, 1988; Newman and Sawusch, 1996; Sawusch and Newman, 2000). If listeners are knowledgeable about within-speaker effects of rate variation on the phonetic dispersion of vowels, wouldn't the same knowledge be applied to across-speaker variations in speaking rate (see Nusbaum and Magnuson, 1997)?

The hypothesis of across-speaker, rate-dependent size of the vowel space was clearly not supported by the current data. The average vowel spaces of the two groups of speakers were virtually identical, as were the average formant fre-

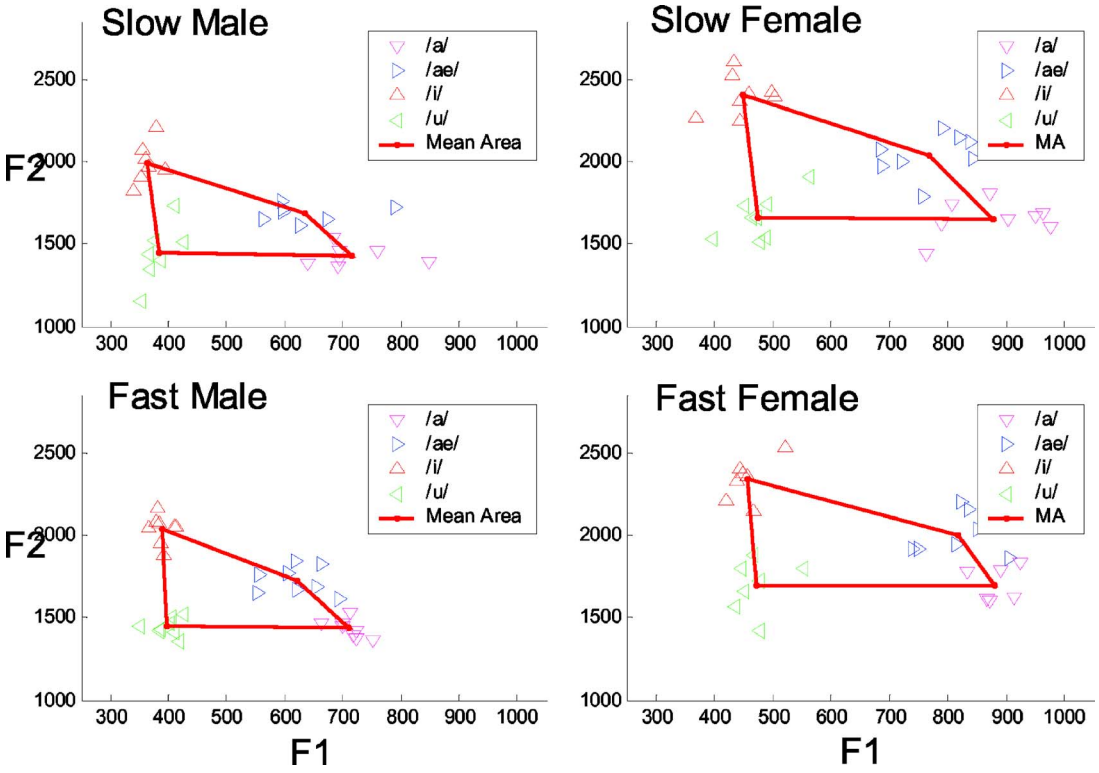


FIG. 1. (Color online) Vowel quadrilaterals plotted by talker groups (slow vs fast) and by gender. Vowel space area (Hz^2) may be inferred from the lines connecting symbols representing mean values of F1 and F2 for vowels /i/, /æ/, /u/, and /a/.

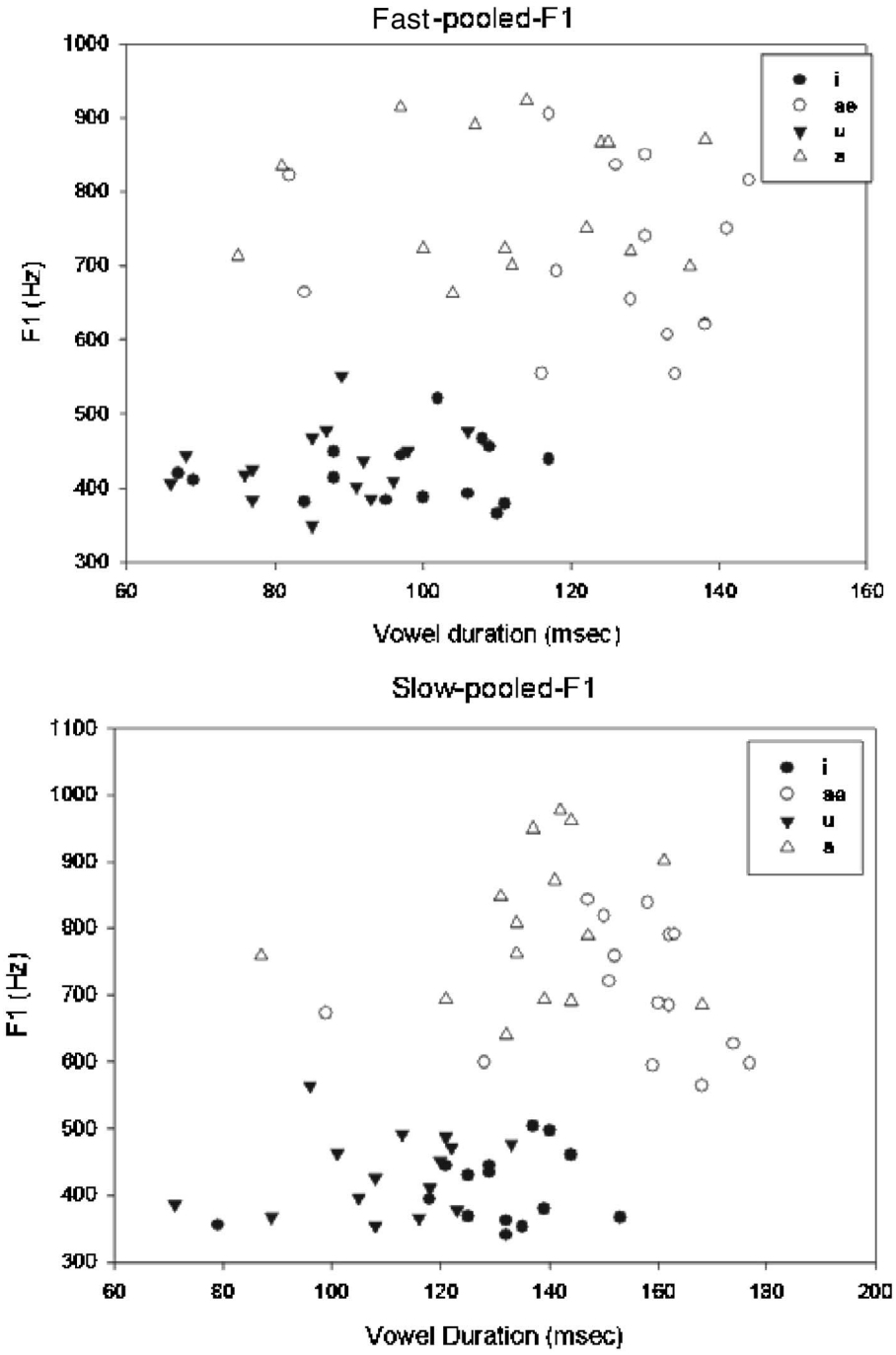


FIG. 2. Plots of F1 as a function of vowel duration for the four corner vowels, pooled across talkers; plots are for the only significant functions between formant frequency and vowel duration. Vowels are coded by separate symbols, and filled and unfilled points plot data for high and low vowels, respectively.

quencies for the individual vowels. Moreover, there was no evidence of a systematic, across-speaker relationship between vowel duration and either F1 or F2 values. It seems reasonable to conclude, therefore, that when vowel acoustics are represented by formant frequencies extracted from the temporal middle of the vowel the habitual rate of a particular speaker should not be expected to have an influence on the typical size of the vowel space. This might suggest that the vowel space—at least its extreme limits—is a rather stable phonetic characteristic of English even in the face of substantial talker variability in the dimension of speaking rate and hence vowel duration.

An alternate view is that vowel-space compression is not a necessary outcome of increased speaking rate either within or across talkers, and that the hypothesis of a relationship between intertalker variation in habitual rate and size of the vowel space may be oversimplified. As noted in the Introduction, van Son and Pols (1990) failed to find formant reduction at a fast, as compared to slower rate, and concluded that rate change may involve adjustments in speaking style that affect vowel durations but not formant frequencies. Clearly, a speaker can override what may be "natural" effects of speaking rate modification—such as vowel reduction—and maintain or enhance acoustic distinctions even at fast rates if given instructions to that effect (see Krause and Braida, 2002) or if trained to do so (as may be likely in van Son and Pols, 1990). In the current experiment, however, none of the speakers had professional voice training and all

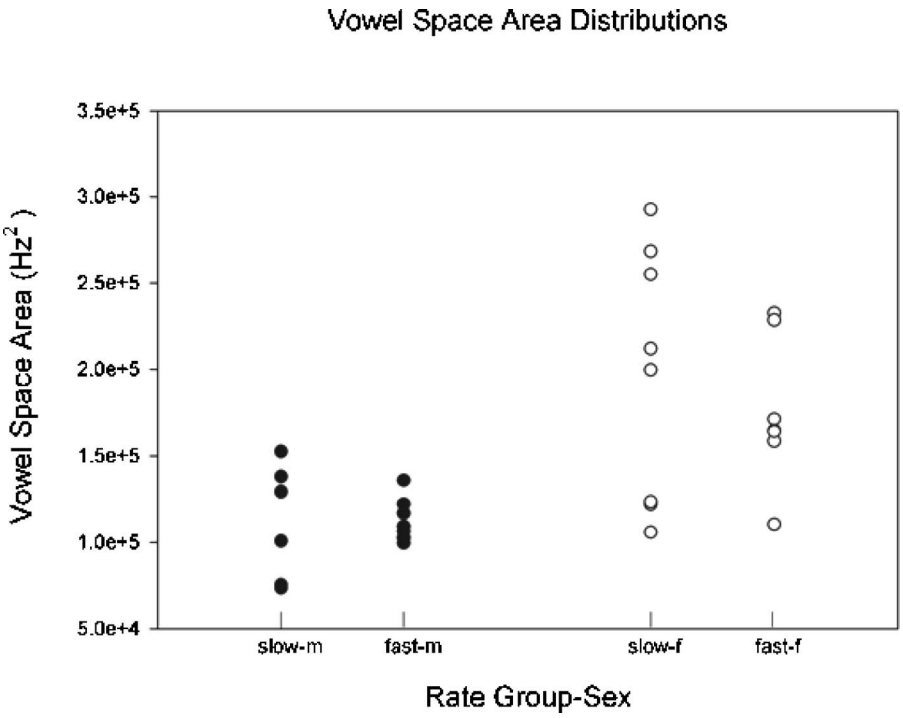


FIG. 3. Individual talker vowel space areas plotted for each of four rate-sex groups; the tendency for larger vowel spaces for female talkers (unfilled symbols) is easily seen, as is the greater spread of points for slow, as compared to fast talkers.

produced the connected speech sample at their self-selected, habitual rates. We consider it unlikely that talkers in the fast group were using a different production "style" from slow talkers to counteract a natural, rate-related tendency for vowel reduction.

It is also possible that vowel production differences exist between these habitually slow and fast talkers, but were not revealed by the simple measurement approach which involved an estimate of formant frequencies over a single, brief interval spanning the temporal midpoint of the vowel. To address this question, studies are needed in which formant frequency histories for vowels are obtained for different speaking rates. van Son and Pols (1992), one of the only studies to examine this kind of effect, failed to find the kind of rate-induced shape changes in first formant history that would be consistent with this possibility. It should be reemphasized that the several demonstrations of speaking rate effects on the size of the acoustic vowel space have used the single-window measurement approach.

Theories of speech production and perception that attempt to account for speaker variability (e.g., Johnson, Ladefoged, and Lindau, 1993; Westbury, Hashi, and Lindstrom, 1998) presumably require these kinds of data, wherein relative variability and stability of different variables are revealed. Perhaps the apparent stability of F-patterns across speakers with very different rates is not all that surprising, although it does raise questions about the underlying articulatory strategies that achieve the same output on different time scales. A related question is whether the results reported here would hold in different languages, and especially those having relatively sparse vowel spaces where the demands for articulatory precision may not be as great as in English which has a larger number of vowels than most languages of the world (see, for example, Manuel, 1990; and Schwartz, Boë, Vallée, and Abry, 1997). Even in English, however, the vowel spaces across rate groups could have been systematically different without (presumably) affecting the perceptual categorization of the vowels defining the space.

In fact, there was variability in the size of the vowel space across talkers within any of the four rate-sex groups, but of more interest is what appears to be a systematic difference in intertalker variability of the vowel space across rate groups. The virtually identical vowel space areas across rate groups were accompanied by clearly larger inter-talker variabilities for slow, as compared to fast talkers (Table IV). Individual talker vowel space areas are plotted in Fig. 3 for each of the four groups; the expected tendency for larger vowel spaces for female talkers (unfilled symbols) is easily seen, as is the greater spread of points for slow, as compared to fast talkers. Because this difference in variability is seen for both sexes, the observation is unlikely to be explained as an artifact of sampling error. Moreover, the more tightly clustered vowel spaces of fast talkers are not clearly on the low side of the distribution of slow-talker vowel spaces, as might be expected from the within-speaker effect of rate on vowel space area. Examination of formant frequency data in Table III shows that for seven of the eight measured formant frequencies, the slow talkers had larger average variabilities than the fast talkers, but such variability in individual formants does not preordain the differences in vowel space variability plotted in Fig. 3. It appears, then, that there is a difference between slow and fast talkers in vowel space area, one that could have implications for calibration and training of automatic speech recognition devices as well as theories of speech perception in which vowel-space normalization is assumed to precede segment- or syllable-level identification. We would therefore revise the tentative statement made above concerning the relative across-talker stability of the

vowel space area (compared to speaking rate) by noting that such stability may depend on which end of the rate continuum the talker belongs to.

Why would the vowel space areas of slow talkers be more variable as a group than those of fast talkers? The standard deviation data in Table I do not support the idea that the group variability difference in vowel space was tied to a corresponding group variability difference in vowel durations. It is possible, however, that in complex movement tasks solutions for a particular goal—such as the shaping of the vocal tract for a particular vowel—are more variable across individuals at slower, as compared to faster rates of action. Slow, as compared to fast talkers, may achieve a vowel configuration with a larger variety of articulatory motions, the result being larger intertalker variations in vocal tract configuration even when the end-product F-pattern is perceptually acceptable. Such rate-based differences in the number of solutions available to complete a task have been reported in the motor control literature for complex turn-and-reach movements (Pigeon, Bortolami, DiZio, and Lackner, 2003); a parallel study in speech production could provide insight to the underlying nature of speech motor control and the range of movement solutions available for articulation of a particular vowel.

ACKNOWLEDGMENTS

This work was supported, in part, by a NIDCD Grant No. R01 DC00319 and NIDCD Grant No. R01 DC03273, as well as a Junior Faculty Grant from the Grants and Contracts Office at the California State University-Fullerton, CA. We thank Greg Turner for his valuable input on initial drafts of the manuscript, and the staff of the Speech Acoustics Lab at UW-Madison for making the measurements. We also thank two anonymous reviewers for their helpful comments on an earlier version of the manuscript. Portions of this work were presented in "Does interspeaker variation in habitual speaking rate predict the size of the acoustic vowel space?," 137th meeting of the Acoustical Society of America, Berlin, Germany, Spring, 1999.

Crystal, T. H., and House, A. S. (**1982**). "Segmental durations in connected speech signals: Preliminary results," J. Acoust. Soc. Am. **72**, 705–716.

Engstrand, O. (**1988**). "Articulatory correlates of stress and speaking rate in Swedish VCV utterances," J. Acoust. Soc. Am. **83**, 1863–1875.

Fourakis, M. (**1991**). "Tempo, stress, and vowel reduction in American English," J. Acoust. Soc. Am. **90**, 1816–1827.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Gay, T. (**1978**). "Effect of speaking rate on vowel formant movements," J. Acoust. Soc. Am. **63**, 223–230.

Gopal, H. S., and Syrdal, A. K. (1988). "Effects of speaking rate on temporal and spectral characteristics of American English vowels, Speech communications," Group Working Papers VI, Research Laboratory of Electronics, MIT, pp. 162–180.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

Hirata, Y., and Tsukada, K. (**2004**). "The effects of speaking rates and vowel length on formant movements in Japanese," in *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception*, Somerville, edited by A. Agwuele, W. Warren, and S. H. Park (Cascadilla Proceedings Project, MA), pp. 73–85.

Johnson, K., Ladefoged, P., and Lindau, M. (**1993**). "Individual differences in vowel production," J. Acoust. Soc. Am. **94**, 701–714.

Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.

Krause, J. C., and Braida, L. C. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Kuehn, D. P., and Moll, K. L. (**1976**). "A cineradiographic study of VC and CV articulatory velocities," J. Phonetics **4**, 303–320.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge, MA).

Lindblom, B. F. F. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **36**, 1773–1781.

Lindblom, B. F. F. (**1990**). "Explaining phonetic variation; A sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. Hardcastle and A. Marchal, pp. 403–439.

Manuel, S. (**1990**). "The role of contrast in limiting vowel-to-vowel coarticulation in different languages," J. Acoust. Soc. Am. **88**, 1286–1298.

Mathworks, Inc. (2000). Using Mathlab version 6 (The Mathworks, Inc. Natick, MA).

Milenkovic, P. (1994). CSPEECH (Computer software) (Department of Electrical and Computer Engineering, University of Wisconsin-Madison, Madison, WI).

Miller, J. L. (**1981**). "Effects of speaking rate on segmental distinctions," in *Perspectives on The Study of Speech*, edited by P. Eimas and J. L. Miller (Lawrence Erlbaum, Hillsdale, NJ), pp. 39–73.

Miller, J. L., and Dexter, E. R. (**1988**). "Effects of speaking rate and lexical status on phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **14**, 369–378.

Miller, J. L., Grosjean, F., and Lomanto, C. (**1984**). "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," Phonetica **41**, 215–225.

Miller, J. L., O'Rourke, T. B., and Volaitis, L. E. (**1997**). "Internal structure of phonetic categories. Effects of speaking rate," Phonetica **54**, 121–137.

Miller, J. L., and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category," Percept. Psychophys. **46**, 505–512.

Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.

Newman, R. S., and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Nusbaum, H. C., and Magnuson, J. S. (**1997**). "Talker normalization: Phonetic constancy as a cognitive process," in *Talker Variability in Speech Processing*, edited by K. Johnson, and J. W. Mullennix (Academic, New York), pp. 109–132.

Ostry, D. J., and Munhall, K. G. (**1985**). "Control of rate and duration of speech movements," J. Acoust. Soc. Am. **77**, 640–848.

Perkell, J. S., and Klatt, D. (**1986**). *Invariance and Variability of Speech Processes* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Pigeon, P., Bortolami, S. B., DiZio, P., and Lackner, J. R. (**2003**). "Coordinated turn-and-reach movements. II. Planning in an external frame of reference," J. Neurophysiol. **89**, 290–303.

Sawusch, J. R., and Newman, R. S., (**2000**). "Perceptual normalization for speaking rate II: Effects of signal discontinuities," Percept. Psychophys. **62**, 285–300.

Schwartz, J-L., Boë, L-J., Vallée, N., and Abry, C. (**1997**). "Major trends in vowel system inventories," J. Phonetics **25**, 233–253.

Simpson, A. P. (**2001**). "Dynamic consequences of differences in male and female vocal tract dimensions," J. Acoust. Soc. Am. **109**, 2153–2164.

Smith, B. L. (**2002**). "Effects of speaking rate on temporal patterns of English," Phonetica **59**, 232–244.

Stetson, R. H. (**1951**). *Motor Phonetics: A Study of Speech Movements in Action* (North-Holland, Amsterdam).

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Tiffany, W. R. (**1959**). "Nonrandom sources of variation in vowel quality," J. Speech Hear. Res. **2**, 305–317.

Tsao, Y.-C., and Weismer, G. (**1997**). "Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component," J. Speech Lang. Hear. Res. **40**, 858–866.

Tsukada, K., and Hirata, Y. (**2003**). "Factors affecting formant frequencies of Japanese vowels: Speaking rates and vowel length," in *Proceedings of the 6th International Seminar on Speech Production*, pp. 296–301.

Tuller, B., Harris, K. S., and Kelso, J. A. S. (**1982**). "Stress and rate: Differential transformations of articulation," J. Acoust. Soc. Am. **71**, 1534–1543.

Turner, G. S., and Weismer, G. (**1993**). "Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis," J. Speech Hear. Res. **36**, 1134–1144.

Turner, G. S., Tjaden, K., and Weismer, G. (**1995**). "The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis," J. Speech Hear. Res. **38**, 1001–1003.

van Bergem, D. R. (**1993**). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," Speech Commun. **12**, 1–23.

van Son, R. J. J. H., and Pols, L. C. W. (**1990**). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **88**, 1683–1693.

van Son, R. J. J. H., and Pols, L. C. W. (**1992**). "Formant movements of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **92**, 121–127.

Weismer, G. (**1984**). "Acoustic descriptions of dysarthric speech: Perceptual correlates and physiological inferences," in *Seminars in Speech and Language*, edited by J. C. Rosenbeck (Thieme-Stratton, New York), Vol. **5**, pp. 293–314.

Weismer, G., Kent, R. D., Hodge, M., and Martin, R. (**1988**). "The acoustic signature for intelligibility test words," J. Acoust. Soc. Am. **84**, 1281–1291.

Weismer, G., Laures, J. S., Jeng, J.-Y., Kent, R. D., and Kent, J. F. (**2000**). "Effect of speaking rate manipulations of acoustic and perceptual aspects of the dysarthria in Amyotrophic Lateral Sclerosis," Folia Phoniatr Logop **52**, 201–219.

Westbury, J. R., Hashi, M., and Lindstrom, M. L. (**1998**). "Differences among speakers in lingual articulation for American English /r/," Speech Commun. **26**, 203–226.

Xue, S. A., and Hao, G. J. (**2003**). "Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study," J. Speech Lang. Hear. Res. **46**, 689–701.

Measuring the critical band for speech[a)]

Eric W. Healy[b)]
Department of Communication Sciences and Disorders, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 29208 and Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-0102

(Received 25 June 2003; revised 15 November 2005; accepted 26 November 2005)

The current experiments were designed to measure the frequency resolution employed by listeners during the perception of everyday sentences. Speech bands having nearly vertical filter slopes and narrow bandwidths were sharply partitioned into various numbers of equal log- or ERB_N-width subbands. The temporal envelope from each partition was used to amplitude modulate a corresponding band of low-noise noise, and the modulated carriers were combined and presented to normal-hearing listeners. Intelligibility increased and reached asymptote as the number of partitions increased. In the mid- and high-frequency regions of the speech spectrum, the partition bandwidth corresponding to asymptotic performance matched current estimates of psychophysical tuning across a number of conditions. These results indicate that, in these regions, the critical band for speech matches the critical band measured using traditional psychoacoustic methods and nonspeech stimuli. However, in the low-frequency region, partition bandwidths at asymptote were somewhat narrower than would be predicted based upon psychophysical tuning. It is concluded that, overall, current estimates of psychophysical tuning represent reasonably well the ability of listeners to extract spectral detail from running speech.
 [DOI: 10.1121/1.2162176]

PACS number(s): 43.71.Es, 43.71.An, 43.66.Fe [PFA] Pages: 1083–1091

I. INTRODUCTION

The division of the auditory spectrum into a series of critical bands by auditory filters is a primary stage of auditory processing. The importance of the critical band (CB) is underscored by its relevance to a wide range of auditory phenomena, including masking and loudness. The psychophysical techniques currently used to measure the critical band (e.g., Patterson, 1976; Patterson and Moore, 1986; Glasberg and Moore, 1990) employ simple tone and noise stimuli, which allow precise control over the input stimulus. However, accurate estimates of frequency tuning become more difficult to obtain as stimulus conditions are made more complex.

An important question is whether the CB measured using relatively simple stimuli represents the frequency resolution underlying the perception of complex sounds such as speech. One technique that has been employed to examine the normal frequency resolution employed when processing speech is to present listeners with stimuli having reduced frequency specificity and to examine resulting reductions in performance. However, performance on broadband speech in quiet can be quite good despite extensive spectral smearing (Baer and Moore, 1993; but also see Boothroyd *et al.*, 1997). Only when speech was presented in background noise did performance suffer with spectral smearing beyond the width of a psychophysical CB (e.g., Celmer and Bienvenue, 1987; ter Keurs *et al.*, 1992, 1993; Baer and Moore, 1993).

Indeed, the minimum frequency resolution required for normal-hearing listeners to understand broadband speech is extremely low: Shannon *et al.* (1995) demonstrated high levels of sentence recognition under conditions in which spectral information was reduced to that conveyed by only three or four broad amplitude-modulated (AM) noise bands. In accord with results involving spectral smearing, broadband speech presented in background noise required additional spectral detail and therefore additional AM frequency channels to reach ceiling intelligibility values (Dorman *et al.*, 1998).

Due to the robust nature of speech and the multiplicity of cues available to the listener, the amount of spectral smearing required to reduce broadband speech intelligibility from ceiling values cannot tell us about the normal frequency resolution employed when processing speech, or the maximum spectral resolution listeners can employ to gain useful information from the signal. This is also true when examining the smallest number of AM carrier bands capable of providing intelligibility of broadband speech. Instead, these measures provide useful information concerning robustness or the resistance of speech to spectral degradation. As smearing or spectral reduction becomes more and more severe,

[a)]Portions of this work were presented at the 139th Meeting of the Acoustical Society of America, Atlanta, GA, 30 May–3 June 2000 [J. Acoust. Soc. Am.**107**, 2914 (2000)].
[b)]Electronic mail: ewh@sc.edu

these manipulations provide information concerning the minimum spectral resolution that can provide intelligibility.

In the current experiments, a technique for measuring the frequency resolution employed by listeners when processing everyday sentences is employed. This technique allows the measurement of frequency resolution within a restricted range of frequencies and the establishment of the speech critical band (*S*-CB).

This measurement is based upon several recent and related findings. The first involves the mechanism allowing for the high intelligibility of narrow-band sentences: When spectral information was removed from narrow-band filtered speech, while maintaining temporal information, intelligibility fell from values near 100% to values near 0%. However, when a minimal spectral contrast was reintroduced by partitioning the band into a pair of juxtaposed temporal patterns, some intelligibility returned (Healy and Warren, 2003). Thus, the near-perfect intelligibility observed when sentences are filtered to a narrow "spectral slit" (Warren *et al.*, 1995) is attributable to contrasting temporal patterns of amplitude fluctuation within the narrow band.

The second finding involves the role of the skirts in the intelligibility of filtered speech: When the CID sentences (Davis and Silverman, 1978) were filtered using a fixed passband of 1/3 octave centered at 1500 Hz, but different slopes, mean intelligibility scores fell from values near 100% for slopes of approximately 100 dB/octave, to below 20% for bands created using high-order FIR filters that produced slopes over 1000 dB/octave (Healy, 1998; Warren and Bashford, 1999; Warren *et al.*, 2004). Thus, much of the contrasting pattern information providing for narrow-band sentence intelligibility can reside within the filter skirts, and a detailed examination of spectro-temporal speech information requires extremely precise filtering.

The third finding involves the influence of spectral overlap of contrasting temporal patterns on intelligibility: When a pair of juxtaposed narrow patterns (AM carrier bands) had shallower filter slopes so that they overlapped spectrally, intelligibility was found to be reduced relative to conditions in which steep filtering was employed to eliminate acoustic overlap of the adjacent patterns (Healy and Warren, 2003).

In the current study, a narrow band of speech having extremely steep filter slopes was sharply partitioned into increasing numbers of component bands, and the temporal envelope of each was used to modulate a corresponding carrier band. The AM carriers were summed and presented to listeners. Spectral detail within the band was manipulated by changing the number of component bands, while holding the overall bandwidth constant. Performance was expected to increase as the number of component bands increased and eventually reach asymptote. The amount of spectral detail in the signal at the point of asymptote provides an estimate of frequency resolution employed when processing the speech band, as additional spectral detail is ineffective.

Physiological limitations of the auditory system apply to the processing of all signals. However, the frequency resolution employed during the processing of speech may potentially be governed by a number of factors including those attributable to the auditory system and, in the case of speech, those of the signal. According to the band importance functions of the articulation and speech intelligibility indexes (ANSI, 1986; 1997), speech possesses maximum density of information in the region surrounding approximately 1500 Hz. This spectral density is reflected in higher relative contributions to intelligibility of fixed width (e.g., 1/3 octave) bands. One possible outcome of the current study is that the speech signal will lack sufficient spectral density of information to take full advantage of the resolving power of the auditory system, especially in the low or high regions of the speech spectrum, and resolution will instead be governed by attributes of the signal. Alternatively, resolution may be quite fine and match or even exceed that predicted by psychophysical tuning.

The goal of the current study was to assess the component bandwidth at asymptote, a measure of the *S*-CB, in each frequency region of the speech spectrum. Frequency resolution was measured in the middle speech frequencies in Experiment 1, and in the high and low frequency regions in Experiments 2 and 3. In Experiment 4, the influence of different carrier types was assessed.

II. EXPERIMENT 1: MEASURING RESOLUTION IN THE MIDDLE FREQUENCIES

A. Experiment 1a

A preliminary experiment was performed to determine the relationship between bandwidth and intelligibility for the sentence materials employed. The restriction of speech to a narrow band was required to reveal the maximum resolution of contrasting temporal patterns within the band. This information was then used to guide the selection of overall bandwidths used for measuring resolution in Experiment 1b.

1. Method

a. Subjects. A group of 12 young adult listeners participated and received either course credit or money in compensation. All were native speakers of English between the ages of 18 and 40 years (mean age of all 124 subjects tested=22 years) and had pure-tone audiometric thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz (ANSI, 1996). Care was taken to ensure that none of the listeners had any prior exposure to the sentence materials. These characteristics and compensation procedures were the same for all listeners tested in this study.

b. Stimuli. The stimuli were based upon the standard recordings (20161-Hz sampling, 16-bit resolution) of the Hearing In Noise Test (HINT, Nilsson *et al.*, 1994). The sentences were filtered to a single narrow band having a width of 2/3, 5/6, or 1 octave centered at 1500 Hz. Filtering was performed using a single pass through a 2000-order digital FIR filter implemented in MATLAB. These parameters produced extremely steep filter slopes, which measured well over 1000 dB/octave. A secondary filtering pass having a slightly wider bandwidth corresponding to the points where the near-vertical slopes intersected the noise floor served to further attenuate the noise floor and resulted in increased signal-to-noise ratio (S/N). The level of each sentence in the 1-octave band was scaled to play back at a slow rms peak level of 70 dBA, and the narrower bandwidths were created by filtering this equated band. The processed digital signals

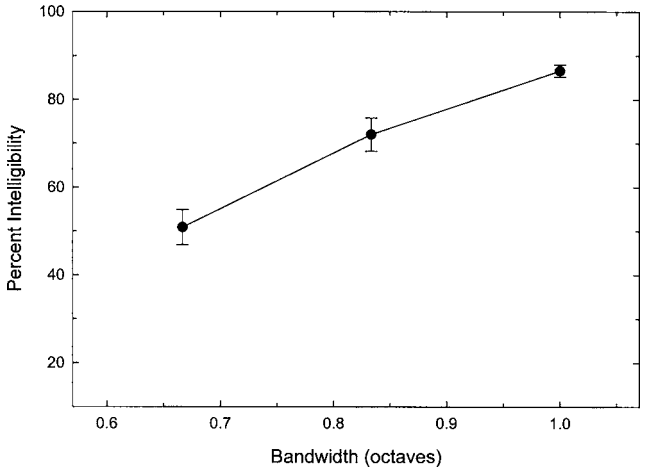


FIG. 1. Group mean intelligibility scores and standard errors for speech bands centered at 1500 Hz and having the bandwidths indicated.

were converted to analog form, amplified (Crown D75), and delivered diotically through TDH-49P headphones mounted in MX/51 cushions.

c. Procedure. Subjects were tested individually, seated with the experimenter in a single-walled audiometric booth located within an acoustically-treated room. Each listener heard 50 test sentences at each of the three bandwidths. The order in which conditions were heard was balanced across subjects so that each appeared in each serial position twice. In addition, to control for potential differences in the difficulty of the 50-sentence sets, the sentence list-to-condition correspondence was balanced so that each list was heard an equal number of times in each condition. Prior to the first condition, subjects heard a single list of 10 HINT practice sentences first broadband, then filtered in a manner corresponding to the first-heard condition. The practice list was repeated in corresponding filtered form before each of the two subsequent test conditions. Subjects were instructed to repeat each sentence aloud after hearing it. They heard each sentence only once, received no feedback, and were encouraged to guess if unsure of the content. The experimenter controlled the presentation of sentences and scored the proportion of component words reported correctly.

2. Results

Figure 1 shows the group mean intelligibility scores and standard errors for the three speech bandwidths presented in this experiment. Intelligibility increased with increasing bandwidth, reaching 87% at 1 octave. It is important in the subsequent experiment to employ a bandwidth at or near the value sufficient for intelligibility, but below ceiling values. Therefore, the data of Experiment 1a provide a guide for the selection of overall bandwidths for further examination in Experiment 1b.

B. Experiment 1b

In this experiment, spectral information within the narrow speech band was quantized by partitioning the band, and removing spectral information from each partition by replacing it with a carrier band that was amplitude modulated by the envelope of the corresponding speech partition.

1. Method

a. Subjects. A total of 30 listeners participated using selection and compensation procedures employed previously.

b. Stimuli. Because it provides appreciable intelligibility, but yields scores below ceiling values, the 1-octave speech band was first selected for the measurement of frequency resolution. In addition to that band, four additional conditions were prepared by partitioning the band into 2, 4, 6, and 10 equal log-width subbands. These subbands were contiguous, meeting at the −6 dB cutoffs. The lowest and highest partitions were created using a low pass or high pass, and the inner bands were created using a bandpass. The FIR filter order for this processing was increased to 6000, to further enhance the acoustic isolation of the juxtaposed bands. Because the FIR filter is linear in phase, all component bands were exactly aligned in time. The partitioning of the speech band is shown in Fig. 2.

Low-noise noise (LNN), which is noise engineered to have extremely small fluctuations in amplitude (Pumplin, 1985; Hartmann and Pumplin, 1988), was selected for the carrier signal. Kohlrausch *et al.* (1997) have described convenient methods for generating LNN, one of which ("Method 1") was used in the current study. This method involves the division of the waveform by its envelope in a series of iterations (100 in this case). Low-noise noise carriers were selected over Gaussian noise carriers because the random amplitude fluctuations of the narrow-band noises could potentially dilute the temporal details of the speech. They were selected over tonal carriers to allow spectral density to remain constant as the number of partition bands changed. Carrier bands having the same frequency composition as the speech partitions were created by summing sinusoidal components having appropriate amplitude and phase, and 0.5-Hz spacing. This component spacing produced a repeated noise having a duration that was sufficiently long (2 s) to not substantially interfere with the perception of the sentences. The LNN carrier bands were separated by 0.5 Hz so that when combined, the entire array would have equal spacing of components.

The amplitude envelope was extracted from each speech partition by full-wave rectification and low-pass filtering (2000-order FIR, 100-Hz cutoff) and applied to a corresponding LNN carrier band by multiplying on a sample point-by-point basis. The AM LNN carriers were then postfiltered to strictly restrict them to the frequency region of origin using the same filters employed to create the speech partitions. Because different filter orders were employed for the inner (6000 order) and outer (2000 order) cutoffs of the lowest- and highest-frequency bands, corrections for the different group delays were performed to ensure exact temporal alignment of the modulated carriers. These manipulations were all implemented in MATLAB.

The AM carriers comprising each condition (1, 2, 4, 6, and 10 bands) were assembled for presentation to listeners. Because this processing preserved the relative overall level of each component band, this resulting array maintained the spectral profile of the original speech band (see Fig. 2 lowest two panels). Each sentence in each condition was presented at a slow rms peak level of 70 dBA using the apparatus employed in the previous experiment. In addition to these conditions based on the 1-octave speech band, additional confirmatory conditions were prepared using bands that were narrower (2/3 octave) and wider (3/2 octave). These overall bandwidths were chosen so that when parti-

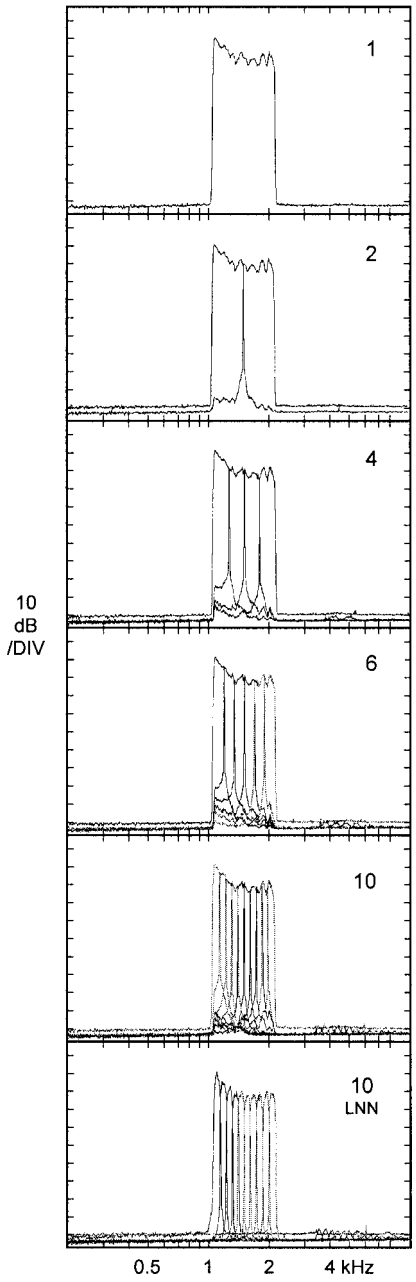


FIG. 2. Overlaid average amplitude spectra for the 1-octave speech band partitioned into 1, 2, 4, 6, and 10 subbands. The bottom panel shows the average amplitude spectra for a corresponding array of 10 amplitude-modulated low-noise noise carrier bands.

tioned into the same numbers of component bands, correspondences in bandwidth would occur. For example, a 1/6-octave component bandwidth is obtained by both dividing the 1-octave band into six and the 2/3-octave band into four. These bands were also centered at 1500 Hz and were presented as arrays of 1, 2, 4, 6, and 10 AM carriers using the same procedures employed to create the 1-octave stimuli.

c. Procedure. Separate groups of 10 listeners each were employed for the three overall bandwidths. They heard 30 test sentences in each of the five spectral resolution conditions. A practice list of 10 sentences was presented in broadband form at the beginning of the session, and was presented first as the overall speech band, then again in a form matching the particular experimental condition prior to each condition. The sentence list-to-condition correspondence was balanced, and the conditions were presented in ascending order (1, 2, 4, 6, and 10 carriers), so that listeners were familiarized with the general procedure during the lower intelligibility conditions (1 and 2 carriers). Following these five AM carrier conditions, listeners heard 50 additional sentences in the overall speech band condition. This allowed performance across the two types of stimuli (AM carrier array versus speech band) to be compared within the same group of listeners. As before, each sentence was played only once and the experimenter scored the proportion of component words correctly recalled.

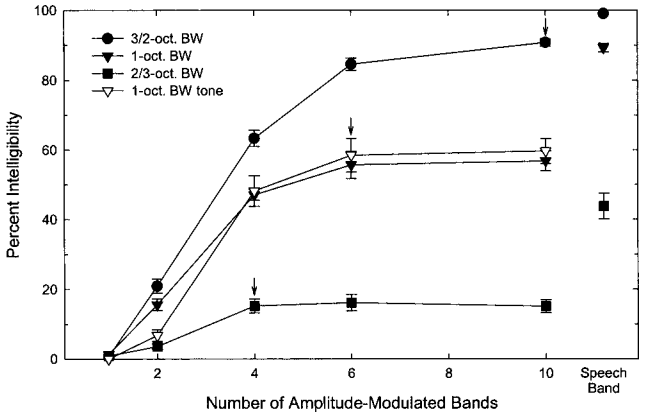


FIG. 3. Group mean intelligibility scores and standard errors for arrays of amplitude-modulated low-noise noise (LNN) carriers having increasing frequency resolution. Shown are scores for overall bandwidths of 3/2, 1, and 2/3 octave all centered at 1500 Hz. Asymptotic performance (indicated by arrows) occurred at 10, 6, and 4 bands respectively, which corresponds to a partition bandwidth of approximately 1/6 octave in each case. At the far right are scores for the corresponding intact speech bands. Also shown (as open symbols) are scores for additional conditions in which pure-tone carriers replaced the LNN carriers.

2. Results

The group mean intelligibilities for the 1-octave overall bandwidth conditions are presented in Fig. 3 as closed inverted triangles. Asymptote was defined throughout by scores differing by less than 2%. As can be seen, performance increased with increasing number of partitions/carriers, with no further increases after six. Also presented is the performance of these listeners on the 1-octave speech band (far right). It was found that performance at asymptote for the AM carriers was below that of the corresponding speech band.

The scores for the confirmatory conditions are also presented in Fig. 3. Performance was found to increase and reach asymptote at four partitions in the 2/3-octave condition (squares). In the 3/2-octave condition (circles), performance continued to increase to the largest number of partitions employed (10), reaching an apparent asymptote at a high level of intelligibility. In each case, the partition bandwidth at asymptote, a measure of the *S*-CB, matches or approximates 1/6 octave.

To further confirm the point of asymptote, the 1-octave conditions were recreated using pure-tone carriers in place of LNN carriers. The tones had a frequency corresponding to the (log) center of each partition. Five additional listeners were recruited and experienced procedures identical to those

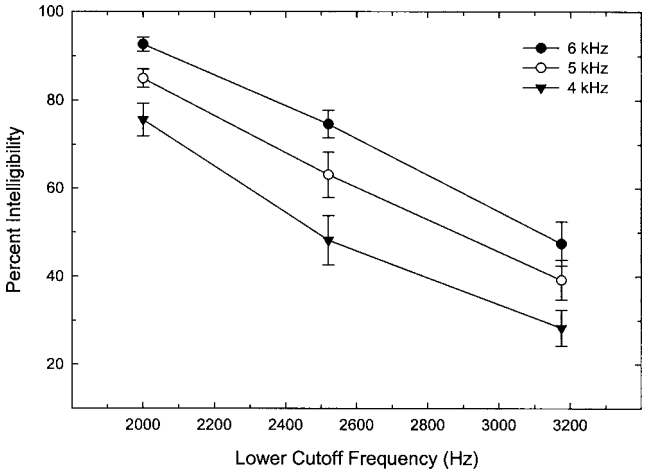


FIG. 4. Group mean intelligibility scores and standard errors for speech bands having lower cutoff frequencies indicated on the abscissa and upper cutoff frequencies as the parameter.

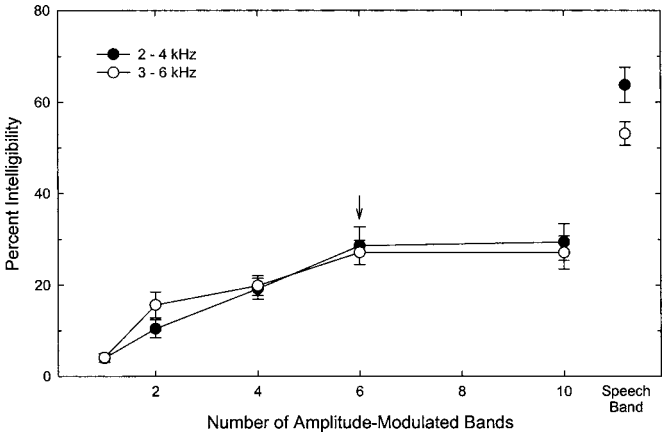


FIG. 5. Group mean intelligibility scores and standard errors for arrays of amplitude-modulated LNN carriers having increasing frequency resolution. Shown are scores for a pair of bands in the high speech frequency region. Asymptotic performance (indicated by the arrow) occurred at 6 bands for both 1-octave bandwidth conditions, corresponding to a partition bandwidth of 1/6 octave. At the far right are scores for the corresponding intact speech bands.

employed earlier, except that the intact speech band was not presented after the modulated carrier arrays. As with the LNN carriers, performance increased and reached asymptote at six bands (see Fig. 3, open symbols). This correspondence between tonal and LNN carriers was observed despite large differences in the spectral density of the tonal carrier array as the number of carriers increased, versus the relative constancy of the LNN arrays.

III. EXPERIMENT 2: MEASURING RESOLUTION AT HIGH FREQUENCIES

A. Experiment 2a

As in Experiment 1, results from a preliminary experiment in which spectral region was related to intelligibility were used to select bands for resolution measurement.

1. Method

A group of nine listeners served. The HINT sentences were filtered using the same primary- and secondary-filtering techniques employed in Experiment 1a. A set of three lower cutoff frequencies (2000, 2520, and 3175 Hz) and three upper cutoff frequencies (4000, 5000, and 6000 Hz) yielded nine speech-band conditions. The sentence list-to-condition correspondence was balanced, and the conditions were presented in a different random order for each listener. Listeners heard 20 sentences in each condition, and each condition was preceded by practice as in Experiment 1a. Each sentence was set to 70 dBA and delivered diotically over Sennheiser HD 250II headphones, which were selected to provide a frequency response that was wider than the audiometric headphones employed in Experiment 1. As before, the experimenter was seated with the subject within an audiometric booth, controlled the presentation of sentences, and recorded the responses.

2. Results

Group mean intelligibility scores are shown in Fig. 4. These results show how intelligibility varies as a function of lower cutoff, and indicate that information up to 6 kHz contributes to intelligibility for these materials under these processing conditions.

B. Experiment 2b

1. Method

Twenty listeners were randomly divided into two groups of ten each. The standard 1-octave bandwidth selected in Experiment 1b was again employed, but the band was transposed upward in frequency to 2–4 kHz in one condition and to 3–6 kHz in a second. The preparation of amplitude-modulated LNN stimuli was the same as that employed in Experiment 1b, and again yielded bands represented by 1, 2, 4, 6, and 10 carriers. The testing procedures were also the same as those of Experiment 1b, with the exception that subjects heard 30, rather than 50 sentences in the overall speech band condition at the end of the session, and that the Sennheiser headphones were employed.

2. Results

The group mean intelligibility scores are shown in Fig. 5. In accord with the results from the mid-frequency region, performance increased and reached asymptote at six bands for both of the 1-octave conditions, corresponding to a component bandwidth of 1/6 octave in both cases.

IV. EXPERIMENT 3: MEASURING RESOLUTION AT LOW FREQUENCIES

A. Experiment 3a

1. Method

The sentences were again filtered using the procedures of Experiment 1a. Four upper cutoff frequencies (476, 566, 673, and 800 Hz) were combined with two lower cutoffs (100 and 200 Hz) to produce eight speech-band conditions.

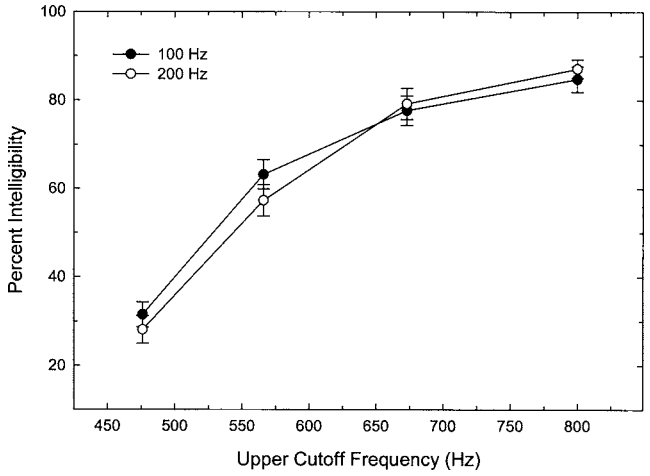


FIG. 6. Group mean intelligibility scores and standard errors for speech bands having the upper cutoff frequencies indicated on the abscissa and the lower cutoff frequencies as the parameter.

Eight listeners heard 30 sentences in each condition, and all procedures were otherwise identical to those of Experiment 2a.

2. Results

Group mean intelligibility scores are shown in Fig. 6. Little difference was observed between the 100 and 200 Hz lower cutoffs, indicating that spectral information below 200 Hz contributes little to intelligibility of these materials under these conditions.

B. Experiment 3b

The partitioning of the speech bands in Experiments 1 and 2 was performed in logarithmic units, because psychophysical tuning, as measured by the equivalent rectangular bandwidth (ERB_N, Glasberg and Moore, 1990; Moore, 2003) follows a simple logarithmic function in the mid- and high-frequency regions, and is approximately constant in width at 1/6 octave. However, in the current experiment, partitioning was performed directly in ERB_N units, because the two functions diverge sharply below approximately 1 kHz. The ERB_N increases to approximately 1/2 octave at 100 Hz, and 1/6-octave bands from 100 to 1000 Hz range from 0.3 to 0.9 ERB_N.

1. Method

A total of 30 listeners were randomly divided into three groups of 10 each. Three overall bandwidths were employed. A four-ERB_N condition spanned the region from 312 to 603 Hz (ERB_N numbers 8–12), a six-ERB_N condition spanned the region from 257 to 698 Hz (ERB_N numbers 7–13), and a 10-ERB_N condition spanned the region from 163 to 921 Hz (ERB_N numbers 5–15). The four- and six-ERB_N bandwidths were divided as before (1, 2, 4, 6, and 10 bands), and the 10-ERB_N bandwidth was divided into 2, 5, 10, 15, and 20 bands. This partitioning was performed in equal ERB_N-width units. Other aspects of the creation of LNN carrier arrays, as well as the apparatus and procedures, were the same as those of Experiment 2b.

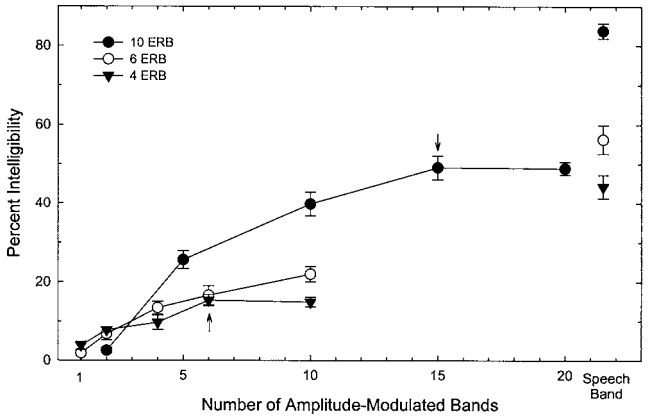


FIG. 7. Group mean intelligibility scores and standard errors for arrays of amplitude-modulated LNN carriers having increasing frequency resolution. Shown are scores for three bands in the low speech frequency region. Asymptotic performance (indicated by arrows) occurred at 6 bands for the 4 ERB_N overall bandwidth condition and at 15 bands for the 10 ERB_N condition, corresponding to a partition bandwidth of 0.67 ERB_N in each case.

2. Results

Group mean intelligibilities are displayed in Fig. 7. The data for the six-ERB_N condition indicates that performance continues to increase beyond six partitions, which suggests that spectral resolution in this region exceeds 1 ERB_N. The data for the four-ERB_N condition also indicate resolution exceeding 1 ERB_N. The asymptote at six bands corresponds to a component bandwidth of 0.67 ERB_N. Finally, the data for the 10-ERB_N condition show an asymptote at 15 bands, which also corresponds to a component bandwidth of 0.67 ERB_N.

V. EXPERIMENT 4: THE INFLUENCE OF DIFFERENT CARRIER TYPES

In the current experiment, differences in performance between the arrays of AM carriers at asymptote and the intact speech bands were examined using different carrier types within a single group of subjects.

A. Method

Ten listeners participated. Performance in one condition (the standard 1-octave overall bandwidth represented by six carriers from Experiment 1b) was compared across four carrier types: pure tone, frequency-modulated (FM) tone, LNN, and Gaussian noise. The pure tones again had frequencies corresponding to the center of each partition. The FM tones had a nominal frequency matching that of the tonal carriers, and had sinusoidal frequency modulation at an average rate of 28 Hz and an average range of 64 Hz. The rate and range values for each FM carrier matched the mean of the dominant frequency modulations of the corresponding LNN carrier (LNN is essentially a flat amplitude, random FM signal; see Kohlrausch *et al.*, 1997). The Gaussian noise band carriers were created using the same filtering used to partition the speech band. These four conditions, plus the 1-octave speech band, were heard in a different random order for each subject. Thirty sentences were heard in each condition, and all

other stimulus preparation methods, apparatus, balancing, and testing procedures were the same as those employed in Experiment 1b.

B. Results

The group mean intelligibility scores (and standard errors) for the five conditions are as follows: speech band: 87.8% (1.7), pure tone: 60.8% (2.8), FM tone: 55.9% (3.1), LNN: 46.5% (2.4), and Gaussian noise: 36.3% (2.9). Thus, none of the carrier arrays reached the intelligibility of the 1-octave speech band.

Performance was poorest with the Gaussian noise carriers, where random amplitude fluctuations of the narrow-band carrier may potentially dilute the temporal details of the speech. Performance was best in the tonal carrier condition, despite the impoverished amplitude spectrum of the 6-tone array and the tonal timbre of this signal. Scores were only slightly reduced when sinusoidal FM was applied to the tonal carriers. Finally, performance in the LNN-carrier condition was below that of the tones, but above that of the Gaussian noise.

VI. DISCUSSION

Measures of auditory frequency resolution have typically employed simple stimuli as well as procedures not compatible with spectro-temporally complex speech signals. Using techniques to quantize spectral information within the speech signal, the current experiments were designed to measure frequency resolution of everyday speech materials in each spectral region separately. This region-specific information is not available when spectral smearing or reduction is applied to all frequency regions simultaneously, or from experiments examining the relation between auditory tuning and broadband speech.

The *S*-CB has been defined here as the width of spectral bands comprising the acoustic speech signal at performance asymptote. This value gives the spectral resolution of the signal beyond which information content does not increase, and reflects the ability to extract spectral detail from the acoustic speech signal. It was important that narrow bandwidths were employed in the current study: If the bandwidths far exceeded required for full intelligibility, then the opportunity to observe the processing of contrasting patterns within the band at maximum resolution would have been lost. For example, if Experiment 1b were repeated with an overall speech bandwidth of two or three octaves, scores would reach ceiling before reaching asymptote. These wider bandwidth conditions would eventually approach those of Shannon *et al.* (1995). Similarly, in the low- and high-frequency regions, it was important to only examine regions of the spectrum that contributed to intelligibility.

When spectral information within the speech band was quantized, scores increased with increasing spectral resolution and eventually reached asymptote, presumably when the spectral resolution of the stimulus matched that of the speech processing system. In the middle speech frequencies (Experiment 1), the component bandwidth at asymptote was found to be 1/6 octave across four conditions employing four different groups of subjects. This correspondence occurred despite differences in overall bandwidth, differences in the number of AM carriers in the array, differences in the level of performance at asymptote, and differences in carrier signal type. It therefore appears that the measurement of the *S*-CB is somewhat robust. In Experiment 2, the *S*-CB was also found to be 1/6 octave in width for a pair of conditions covering the higher regions of the speech spectrum. Because 1/6-octave bands in the region from 1000 to 6000 Hz range from 0.9 to 1.0 ERB_N, it is concluded that, in this frequency region, the size of the *S*-CB approximately matches that of the psychophysical CB. In Experiment 3, resolution was examined in the lower speech frequency region. In contrast to Experiments 1 and 2, it was found that the *S*-CB was somewhat narrower than the psychophysical CB. Again, despite differences in bandwidth, mean intelligibility level, and number of component bands at asymptote, the conditions converged to yield a measurement of the *S*-CB at 0.67 ERB_N. It is unclear why the two measures of frequency resolution diverge somewhat at low frequencies. However, the *S*-CB is within approximately 25 Hz of the CB at a frequency of 500 Hz and so it may be concluded that, overall, current estimates of psychophysical tuning also describe reasonably well the ability of normal-hearing listeners to extract spectral detail from the spectro-temporally complex speech signal.

Because high-context sentences were employed, the current results are restricted to these materials and it is possible that different results will be obtained for individual words, syllables, or even phonemes. However, the goal of the current study was to assess resolution of everyday speech, and the results indicate that the functional density of information in this material is quite high. It is interesting to note that the observed match between speech resolution and psychophysical tuning in the mid and high frequencies, but the modest divergence in the low frequencies would not be predicted from the shape of the band-importance functions of the Articulation or Speech Intelligibility Indexes.

The intelligibility of the array of AM carriers at asymptote does not reach that of the intact speech band. This difference in performance was observed both across and within subjects. It appears that a speech band having spectral information quantized in the fashion employed here does not completely retain characteristics of the signal required for full intelligibility. This is especially evident from conditions in which spectral resolution far exceeds that required for asymptotic performance (e.g., 2/3-octave band represented by 10 carriers in Fig. 3).

A control condition was employed to ensure that this difference in performance between the AM carriers and the intact speech band was not due to some disruption caused by the severe filtering used to partition the speech bands. Two of the subjects from Experiment 1b participated in additional testing at the end of the session. They heard six lists of ten HINT sentences each, alternating between the 1-octave speech band and the 1-octave speech band divided into six partitions and recombined. The average intelligibility score for the two conditions differed by less than 1% (89.9% intact

speech band, 90.5% recombined speech partitions), suggesting that artifacts associated with the extreme filtering did not hinder performance.

A second control condition was designed to investigate possible interference produced by beating of the juxtaposed carrier bands. This test was based on modulation detection interference (Yost and Sheft, 1989). If beating of the carrier array was interfering with the temporal details of the speech, then the introduction of an array of carriers to a separate frequency region may have some perceptible influence on a speech band. An array of six unmodulated LNN bands mimicking those employed as carriers in the 1-octave overall bandwidth condition of Experiment 1 was transposed up in frequency by 2000 Hz and presented along with the 1-octave speech band at an equal average spectrum level. An additional condition employed an array of tones having frequencies corresponding to the centers of the transposed LNNs. Both informal listening and formal intelligibility testing indicated no interference.

Previous work has shown that filter skirts as steep as 100–200 dB/octave can contribute considerably to narrowband speech intelligibility. However, it is difficult to know precisely how far down the filter skirts the fluctuating information is available, and it is difficult to know to what extent the severe spectral tilt of information lying along the slope can affect recognition. In addition, it has been found that intelligibility of narrow juxtaposed temporal speech patterns can suffer when they possess filter skirts that cause them to overlap spectrally. The use of extremely steep filter slopes allows the elimination of these complicating effects.

The current results have implications not only for the processing of speech by normal-hearing listeners, but also for speech perception by individuals having a hearing impairment (HI). Sensorineural HI is often characterized by broadened auditory tuning (for a review, see Moore, 1998). Although it may be assumed that tuning that exceeds the normal psychophysical CB will negatively impact speech perception, the empirical evidence has been mixed (e.g., Celmer and Bienvenue, 1987; Moore and Glasberg, 1987; Thibodeau and van Tasell, 1987; Dubno and Dirks, 1989; Turner and Henn, 1989; Dubno and Schaefer, 1992; ter Keurs *et al.*, 1993; Ching *et al.*, 1997). Because the ability to extract spectral information was found to be quite fine in the current investigation, and to match the psychophysical tuning of normal-hearing listeners in the high frequency region where most hearing loss occurs, these data indicate that any broadening beyond normal may be expected to hinder the ability to extract normally usable information from the speech signal.

The frequency resolution of speech is also a question of importance for the prosthetic treatment of hearing loss. Modern cochlear implants employ an array of electrodes that stimulate different portions of the cochlea with temporal information derived from corresponding regions of the speech spectrum. Although a current goal of cochlear implant development is enhance the poor frequency representation of these devices and provide as much spectral specificity as possible, mimicking the frequency resolution employed normally when processing speech can be considered an ultimate goal of cochlear prosthetics.

VII. SUMMARY AND CONCLUSIONS

When spectral information was quantized within speech bands at or near the minimum bandwidth required for intelligibility, performance was found to asymptote consistently across a variety of conditions at a partition bandwidth matching or approximating 1/6 octave for frequency regions from approximately 1000 to 6000 Hz. Thus, in these regions, the size of the *S*-CB approximately matches that of the psychophysical CB. However, in the region below approximately 1000 Hz, the ability to process spectral contrasts within the running speech signal is somewhat finer than would be suggested by psychophysical tuning.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant Nos. DC01376 and DC05795. The authors thank Dave Eddins for providing the code used to create the LNN and Stuart Rosen for helpful comments on an earlier version of this manuscript.

American National Standards Institute. (**1996**). *ANSI-S3.6 (1996), Specifications for Audiometers*. (New York).

American National Standards Institute. (**1986**). *ANSI-S3.5, 1969 (R 1986), American National Standard Methods for the Calculation of the Articulation Index* (New York).

American National Standards Institute. (**1997**). *ANSI-S3.5. (1997), American National Standard Methods for the Calculation of the Speech Intelligibility Index*. (New York).

Baer, T., and Moore, B. C. J. (**1993**). "Effect of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1997**). "Simulation of sensorineural hearing loss: Reducing frequency resolution by uniform spectral smearing," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ).

Celmer, R. D., and Bienvenue, G. R. (**1987**). "Critical bands in the perception of speech signals by normal and sensorineural hearing loss listeners," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht).

Ching, T., Dillon, H., and Byrne, D. (**1997**). "Prediction of speech recognition from audibility and psychoacoustic abilities of hearing-impaired listeners," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ).

Davis, H., and Silverman, S. R. (**1978**). *Hearing and Deafness*, 4th ed. (Holt, Rinehart, and Winston, New York).

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.

Dubno, J. R., and Dirks, D. D. (**1989**). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1666–1675.

Dubno, J. R., and Schaefer, A. B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110–2121.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Hartmann, W. M., and Pumplin, J. (**1988**). "Noise power fluctuations and the masking of sine signals," J. Acoust. Soc. Am. **83**, 2277–2289.

Healy, E. W. (**1998**). "A minimum spectral contrast rule for speech recognition: Intelligibility based upon contrasting pairs of narrow-band amplitude patterns," Doctoral dissertation, University of Wisconsin-Milwaukee.

Healy, E. W., and Warren, R. M. (**2003**). "The role of contrasting temporal amplitude patterns in the perception of speech," J. Acoust. Soc. Am. **113**,

1676–1688.
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acoustica united with Acoustica **83**, 659–669.
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).
Moore, B. C. J., and Glasberg, B. R. (**1987**). "Relationship between psychophysical abilities and speech perception for subjects with unilateral and bilateral cochlear hearing impairments," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht).
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.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filter shapes and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
Pumplin, J. (**1985**). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
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.
Thibodeau, L. M., and Van Tasell, D. J. (**1987**). "Tone detection and synthetic speech discrimination in band-reject noise by hearing-impaired listeners," J. Acoust. Soc. Am. **82**, 864–873.
Turner, C. W., and Henn, C. C. (**1989**). "The relation between vowel recognition and measures of frequency resolution," J. Speech Hear. Res. **32**, 49–58.
Warren, R. M., and Bashford, Jr., J. A. (**1999**). "Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband," J. Acoust. Soc. Am. **106**, L47–L52.
Warren, R. M., Bashford, Jr., J. A., and Lenz, P. W. (**2004**). "Intelligibility of bandpass filtered speech: Steepness of slopes required to eliminate transition band contributions," J. Acoust. Soc. Am. **115**, 1292–1295.
Warren, R. M., Riener, K. R., Bashford, Jr., J. A., and Brubaker, B. S. (**1995**). "Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits," Percept. Psychophys. **57**, 175–182.
Yost, W. A., and Sheft, S. (**1989**). "Across-critical-band processing of amplitude-modulated tones," J. Acoust. Soc. Am. **85**, 848–857.

On the perception of voicing in syllable-initial plosives in noise[a)]

Jintao Jiang,[b)] Marcia Chen, and Abeer Alwan[c)]
Department of Electrical Engineering, University of California, Los Angeles, California 90095

(Received 16 November 2004; revised 10 November 2005; accepted 15 November 2005)

Previous studies [Lisker, J. Acoust. Soc. Am. **57**, 1547–1551 (1975); Summerfield and Haggard, J. Acoust. Soc. Am. **62**, 435–448 (1977)] have shown that voice onset time (VOT) and the onset frequency of the first formant are important perceptual cues of voicing in syllable-initial plosives. Most prior work, however, has focused on speech perception in quiet environments. The present study seeks to determine which cues are important for the perception of voicing in syllable-initial plosives in the presence of noise. Perceptual experiments were conducted using stimuli consisting of naturally spoken consonant-vowel syllables by four talkers in various levels of additive white Gaussian noise. Plosives sharing the same place of articulation and vowel context (e.g., /pa,ba/) were presented to subjects in two alternate forced choice identification tasks, and a threshold signal-to-noise-ratio (SNR) value (corresponding to the 79% correct classification score) was estimated for each voiced/voiceless pair. The threshold SNR values were then correlated with several acoustic measurements of the speech tokens. Results indicate that the onset frequency of the first formant is critical in perceiving voicing in syllable-initial plosives in additive white Gaussian noise, while the VOT duration is not. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2149841]

PACS number(s): 43.71.Es, 43.71.Bp [ARB] Pages: 1092–1105

I. INTRODUCTION

Research on speech perception and human auditory processes, particularly in the presence of background noise, helps to improve and calibrate such practical applications as noise-robust automatic speech recognition systems (e.g., Hermansky, 1990; Strope and Alwan, 1997) and aids for the hearing impaired. Allen (1994) has argued that the study of speech perception in noise is not only practically important for automatic speech recognition but also critical for a complete and correct understanding of the psychology of human speech perception. The present study examines the contributions of various acoustic parameters to the perception of the feature voicing in syllable-initial plosives in the presence of additive white Gaussian noise. Questions addressed in this paper include the following: How does the perception of voicing vary with signal-to-noise ratio (SNR)? Does the threshold SNR of voicing detection vary with place of articulation and/or vowel context? Which acoustic properties account for the perception of voicing in noise?

The present study focuses on plosives that span three different places of articulation, labials (/b,p/), alveolars (/d,t/), and velars (/g,k/). These six consonants are further classified as voiced {/b/,/d/,/g/} or voiceless {/p/,/t/,/k/}. Plosive consonants are produced by first forming a complete closure in the vocal tract via a constriction at the place of articulation, during which there is either silence or a low-frequency hum (called *voicebar/prevoicing*). The vocal tract is then opened suddenly releasing the pressure built up behind the constriction; this is characterized acoustically by a transient and/or a short-duration noise burst (Stevens, 1998). The period between the release of the plosive and the beginning of voicing in the vowel is called the voice onset time (VOT). During this period there is silence and/or aspiration noise. The time interval between the onset of the following vowel and the instance when a formant frequency reaches its steady-state value is called the formant transition.

Previous studies on the voiced/voiceless distinction in plosives have focused primarily on the VOT duration. It has been shown that the VOT duration of voiced plosives is significantly shorter than voiceless plosives. Liberman *et al.* (1958) conducted perceptual experiments using synthetic syllable-initial plosives (/b/, /d/, and /g/), where the onset of F1 was cut back (delayed) by amounts varying between 10 and 50 ms relative to the burst; F2 and F3, however, began immediately after the release. The authors concluded that the VOT duration (as defined by the amount of F1 cutback) was a cue for voicing, and that replacing F2 and F3 with noise, instead of harmonics, increased the voiceless effect. Lisker and Abramson (1964) measured the VOT duration of several naturally spoken utterances, comprised of {/b/,/d/,/g/,/p/,/t/,/k/} followed by the vowel /a/. The average VOT duration of voiced plosives was found to be 1, 5, and 21 ms for /b/, /d/, and /g/ in English, respectively. In contrast, the average VOT duration of voiceless plosives was found to be 58, 70, and 80 ms for /p/, /t/, and /k/ in English, respectively. There were no values in common between voiced/voiceless pairs of plosives with the same place of articulation (e.g., /b,p/). The VOT duration was therefore a good and easily measurable acoustic property to distinguish between voiced and voiceless plosives. Lisker and Abramson

[a)]Portions of this work were presented at the 139th Meeting of the Acoustical Society of America (2000) and EUROSPEECH 2001. This work is based in part on Marcia Chen's MS thesis, UCLA, 2001.
[b)]Now with the House Ear Institute, Los Angeles, CA 90057.
[c)]Electronic mail: alwan@ee.ucla.edu

(1970) followed up their previous study with perceptual experiments using synthetic stimuli of varying VOT durations. Stimuli were made to sound like {/b,p/,/d,t/,/g,k/}, in English, followed by /a/. Tokens were identified as voiced if their VOT durations were under 15, 20, and 30 ms for /b/, /d/, and /g/, respectively; tokens were identified as voiceless if their VOT durations were over 30, 40, and 50 ms for /p/, /t/, and /k/, respectively. The study concluded that the voicing boundary for English tokens changed along the VOT duration continuum as a function of place of articulation. Lisker and Abramson (1964; 1970) also demonstrated that VOT duration and the voicing boundary along the VOT duration continuum varied from language to language, although there were within-language regularity and cross-language predictability in the realization of voicing contrasts (Cho and Ladefoged, 1999).

More recent research has focused on how different acoustic cues are integrated or traded during perception: Many studies have demonstrated the occurrence of cue interactions (e.g., Fitch *et al.*, 1980; Miller, 1977; Sawusch and Pisoni, 1974) and modeled the interactions (Massaro and Oden, 1980; Repp, 1983). According to Stevens and Klatt (1974; see also Lisker *et al.*, 1977), the VOT duration boundary between voiced and voiceless tokens was unstable and varied depending on the presence or absence of a rapidly changing F1 transition. In the Stevens and Klatt (1974) study, perceptual tests were conducted using synthetic /da,ta/ stimuli with varying F1 transition rates and VOT durations. The authors reported that, for one listener, classification seemed to be based solely on the VOT duration: the VOT duration boundary was always 30 ms regardless of transition slope. For another listener, results seemed to be based on the duration of the F1 transition. The other three listeners seemed to use a mixture of cues. The authors also conducted a separate experiment to show that a transition was perceptually detectable at about 13 ms and a rapid spectrum change was indicative of voicing. Lisker (1975) conducted further experiments on the F1 transition using synthetic stimuli of /ga/ and /ka/ and found that although F1 had a significant effect on the voiced/voiceless classification, it was neither necessary nor as sufficient as the VOT duration. In addition, it was not the dynamic quality (rapidly changing) of F1 but the low F1 onset frequency that indicated voicing. Summerfield and Haggard (1977) refined and extended Lisker's (1975) conclusion. The authors showed that when F1 onset frequency and F1 transition duration were independently controlled in the /g,k/ pair, only F1 onset frequency had a significant effect on voicing perception. When a factorial F1 (four steady-state frequencies) $\times$ F2 (four steady-state frequencies) design was used for the /d,t/ distinction, the authors further ruled out F2 or F1$\times$F2 interaction as possible voicing cues. These results suggest that there may be a perceptual trading relationship between F1 onset frequency and VOT duration (i.e., a low F1 onset frequency, in the context of a high vowel, resulted in a longer threshold VOT) and that the trading relationship is vowel dependent. Hall *et al.* (1995) found that VOT duration and aspiration amplitude were completely integral using speeded classification experiments. Benki (2001) experimented with synthetic vowel-consonant-vowel and consonant-vowel (CV) stimuli, where F1 transition, place of articulation (bilabial, alveolar, and velar), and VOT duration were manipulated. The author found that, in quiet, place of articulation and F1 transition characteristics were both important phonetic cues for voicing categorization and that the magnitude of the F1 transition effects was considerably larger than the place of articulation effect.

Other acoustic cues were also investigated with respect to voicing classification. Peterson and Lehiste (1960) found that the vowel duration was not a reliable cue for voicing in syllable-initial consonants. The loudness of the burst has been speculated to be a distinguishing characteristic for voicing (Lisker and Abramson, 1964): voiceless plosives have a greater articulatory force, or a louder burst, than voiced ones. Klatt (1975) showed that aspiration during the period of closure was present in voiceless plosives, but absent in voiced plosives. In the same study, Klatt reported that the VOT duration changed as a function of the place of articulation and was longer before high vowels than before mid- and low vowels. Repp (1979) showed in perceptual tests with synthetic stimuli that the VOT duration boundary between /da/ and /ta/ was linearly dependent on the ratio between the amplitude of the aspiration and that of the vowel. A one-decibel increase of the ratio led to a shortening of the VOT duration boundary by, on average, 0.43 ms. In the same study, other experiments were conducted with varying burst and aspiration amplitudes. Both acoustic properties were seen to affect the voicing boundary (a louder burst or louder noise aspiration increased the voiceless effect), although the amplitude of aspiration had a larger effect. The fundamental frequency (F0) has also been suggested as a possible cue for voicing in stops (Haggard *et al.*, 1970; Ohde, 1984; Whalen *et al.*, 1993). Haggard *et al.* (1970) examined the perception of synthetic /bi,pi/ stimuli and showed that utterances with a low rising F0 across the transition region indicated voicing, while a high falling F0 indicated voicelessness. Ohde (1984) examined several naturally spoken CVC tokens by male talkers and found that the absolute value of F0 immediately following the onset of voicing was higher in voiceless stops (average 135 Hz) than voiced stops (average 103 Hz). Also, the drop in F0 between the first glottal pulse and the second pulse was larger for voiceless stops (average 16 Hz) than voiced stops (average 3 Hz).

The studies discussed above were all conducted in quiet environments; however, speech is often heard in the presence of background noise. Therefore, it is of great importance to investigate speech perception in noise. Extensive studies on the perceptual confusions between consonants in the presence of noise have been conducted, most notably by Miller and Nicely (1955). Their study used naturally spoken utterances consisting of one of 16 consonants followed by the vowel /a/, in varying levels of noise and band pass filtering configurations. Their work showed that voicing was much less affected by noise than other features (such as the place of articulation). Voicing was still discriminable at SNRs as low as -12 dB, while place information, in contrast, was difficult to distinguish at SNRs less than +6 dB. The percep-

tual experiments for that study, however, allowed confusions between all consonants, and not just the plosives.

The perception of speech sounds in noise also depends on the noise characteristics. Hant and Alwan (2000) examined the perceptual confusion of synthetic plosive consonants in noise and found that there was a 5 to 10 dB drop in threshold SNRs between the perceptually flat and speech-shaped noise, suggesting that adult native English listeners might be using high frequency cues to discriminate plosives in speech-shaped noise, but that those cues were unavailable in white noise. Nittrouer *et al.* (2003) showed clear differences in adults' perception of consonants in white versus speech-shaped noise, while there was no difference in children's perception. In a recent study of English phoneme confusions in multispeaker babble noise by native and non-native listeners, Cutler *et al.* (2004) showed that although both language background and noise were significant factors in perception, the interaction of these two factors was not significant.

Few studies have examined physical measures that could account for the changes in the perception of features or sounds in the presence of background noise (Hant and Alwan, 2000; Hant and Alwan, 2003; Soli and Arabie, 1979). Soli and Arabie (1979) analyzed the consonant confusion data from Miller and Nicely (1955) and suggested (qualitatively) that consonant confusion data could be better explained by the acoustic properties of the consonants than by phonetic features. In (Hant and Alwan, 2000, 2003), the authors developed a general, time/frequency detection computational model to predict human speech perception in noise. The model predicted well the discrimination of synthetic voiced plosive CV syllables in perceptually flat and speech-shaped noise. Their perceptual experiments and model showed that formant transitions are more perceptually salient than the plosive burst is in noise.

Previous literature has focused mostly on the /Ca/ context. Notable exceptions include Klatt (1975), Summerfield and Haggard (1977), and Hant and Alwan (2003). The present study examines how vowel context (/a/, /i/, or /u/) affects the relationship between the acoustic properties of the speech signal and results from perceptual experiments conducted in the presence of additive white Gaussian noise. First, measurements of several acoustic properties from a set of CV utterances were made (in quiet) and analyzed for possible voicing cues. Second, perceptual experiments were conducted using the speech tokens mixed with varying amounts of background noise. Finally, the acoustic measurements were examined in conjunction with the results from the perceptual experiments to determine which cues could possibly account for the perception of voicing in noise. The hypothesis, implicit in this paper, is that the perception of voicing in plosives in noise is affected by an interaction between SNR on the one hand, and factors that affect the acoustic characteristics of the plosive release on the other hand; such factors include talkers' gender, place of articulation, and vowel context. Specifically, post-hoc correlation analyses will demonstrate that tokens whose voicing is specified only by VOT are well recognized at high but not low SNRs, while tokens whose voicing is specified by both VOT and F1 onset frequency are well recognized at both high and low SNRs. That is, CVs possessing strong voicing cues in the vowels, such as the first formant frequency and VOT, yield better voicing perception in noise than those not having strong F1 onset frequency cues. In addition, we also analyzed the importance of other acoustic cues that have been hypothesized in the literature to be important for voicing classification (such as F0, the burst, and voicebar).

TABLE I. CV pairs used in this study.

Labials	Alveolars	Velars
/b a, p a/	/d a, t a/	/g a, k a/
/bi, pi/	/di, ti/	/gi, ki/
/bu, p u/	/du, t u/	/gu, k u/

II. METHODS

A. Stimuli

The stimuli consisted of isolated, naturally spoken consonant-vowel utterances (CVs), each comprising of a plosive from the set {/b/, /p/, /d/, /t/, /g/, /k/} followed by a vowel from the set {/a/, /i/, /u/}, for a total of 18 syllables. The speech signals were recorded in a sound attenuating room using a headset microphone, and were then sampled (sampling rate of 16 kHz with a 16 bits per sample representation). Four talkers, all native speakers of American English between the ages of 18 and 36 years, were recorded. Two talkers were male, and two were female. Each talker produced eight repetitions of each CV, but only four tokens were used for the current study (the first three tokens and the last one were discarded), resulting in a total of 16 tokens per CV. Syllables were sorted in voiced/voiceless pairs (such as /b a/ and /p a/), such that place of articulation and vowel context were identical, and the two syllables in each pair differed only in the voicing dimension. Thus, there were a total of nine CV pairs (see Table I).

The masking noise used in the perceptual experiments was a 1250 ms segment of white Gaussian noise. At the beginning of each experimental session, 32 Gaussian noise sources were generated; each noise sample was generated randomly and modeled with a Gaussian distribution. During the presentation of each stimulus, a noise masker was randomly selected from the 32 Gaussian noise sources. The SNR was defined as the ratio of the peak root mean square (RMS) value in the CV to the RMS value of the noise token $[20 \log 10(\text{peak\_RMS}_{\text{CV}}) - 20 \log 10(\text{RMS}_{\text{noise}})]$. The first term occurred in the vowel part for most of the CV tokens. The peak RMS energy of a token was computed using a 30 ms rectangular window. The RMS energy of the noise was based on the entire noise segment. Hence, the SNR did not depend on the duration of the speech token.

B. Acoustic measurements

For acoustic measurements, tokens were first decimated to an 8 kHz sampling rate and then pre-emphasized with a coefficient of 0.97. All tokens were normalized such that the

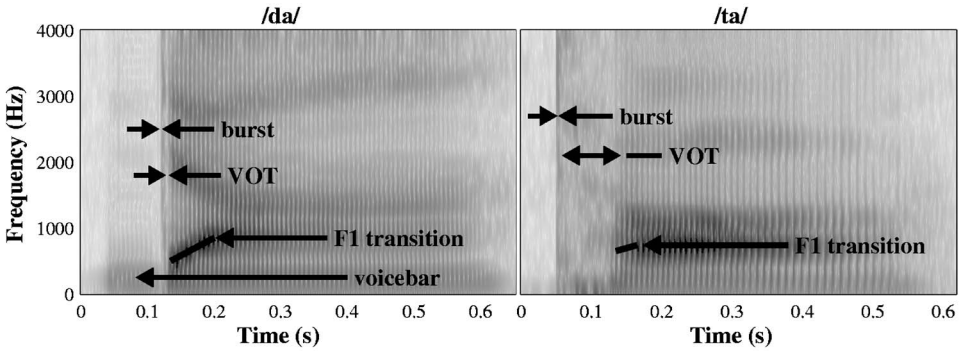


FIG. 1. Spectrograms of /d a/ and /t a/ tokens, sampled at 8 kHz, illustrating different acoustic properties (burst, VOT, F1 transition, and voicebar).

peak amplitude of the entire sampled waveform was set to the same level. Acoustic measurements were made for the speech tokens without noise.

The voicebar, burst, VOT, and F0 measurements were made by visually inspecting the time waveforms and wideband spectrograms of the tokens using the software CoolEdit Pro. Wideband spectrograms were calculated using a 6.4 ms Hamming window with frame shift of one sample (see Fig. 1). The voicebar/prevoicing was defined as the periodic low-frequency energy before the consonant release. The burst was defined as the short segment characterized by a sudden, sharp vertical line in the spectrogram. The segmentation of a burst was performed visually by examining both the waveform and spectrum in CoolEdit Pro. If multiple bursts were present, the burst duration (in ms) was measured from the beginning of the first burst to the end of the last. The maximum absolute amplitude of the sampled waveform in the burst segment was defined as burst peak amplitude (in dB). VOT duration was measured from the end of the burst to the beginning of the vowel, which was also the beginning of the first waveform period. VOT peak amplitude (in dB) was the maximum absolute value of the sampled waveform in the segment. F0 at the onset of the vowel was calculated from the inverse of the first pitch period, measured from peak to peak. The steady-state F0 frequency was similarly calculated from the length of a pulse measured at approximately 100 ms after the onset of the vowel. An F0 frequency change was then derived from the F0 onset and steady-state frequencies (as a drop in F0).

Formant frequency measurements were made from the time waveforms, spectrograms, LPC spectra [Fig. 2(a)], and short-time DFT spectra [Fig. 2(b)] using Matlab. A 20 ms (for tokens from male talkers) or 15 ms (for tokens from female talkers) Hamming window was applied to define an analysis segment. Different Hamming window lengths were used because of the differences in the fundamental frequency between male and female talkers. To obtain a spectrum, each segment was zero-padded for a 1024-point FFT analysis, and the frame shift was half the Hamming window length (i.e., 10 ms for male talkers and 7.5 ms for female talkers). For an LPC analysis, no zero padding was applied, the frame shift was 2.5 ms for all talkers, and the LPC order was between 8 and 12. The vowel was defined to begin when the vocal folds began to vibrate after aspiration. Vowel measurements included the first three formants (F1, F2, and F3). The three formants were located by examining the LPC spectra [Fig. 2(a)], FFT spectra [Fig. 2(b)], and spectrograms. Three landmark points were defined for each formant, formant onset, offset, and steady state [Fig. 2(c)]. The onset of the vowel, chosen manually, was defined as the center point of the frame that exhibited the following characteristics: an abrupt increase in the total energy of the frame and a sudden change in the spectral properties, particularly the introduction of a sharp F1 spectral peak. F1 always began in this frame, but not necessarily F2 or F3. Accordingly, F2 and F3 onsets were defined by examining the sudden spectral change in F2 and F3 frequency range, respectively. The end of a formant transition (offset), chosen automatically, was defined as the frame during which the rate of change of the formant frequency fell to less than 5 Hz per 2.5 ms, and the average rate of change for the next 12.5 ms was also less than 5 Hz per 2.5 ms [Kewley-Port, 1982; see Fig. 2(d)]. The steady-state point was centered at 95 ms after the onset, and the steady-state measurements were averaged over five frames. At the

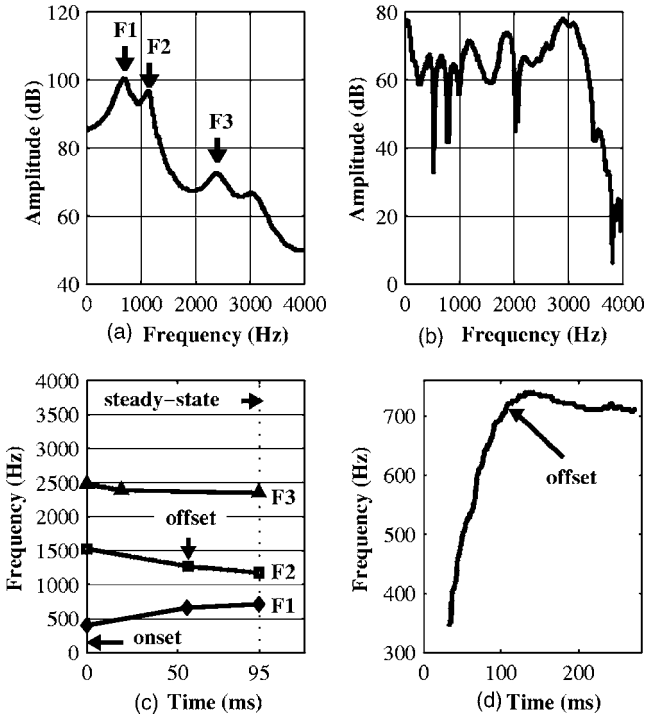


FIG. 2. (a) LPC spectrum of a /t a/ token during the vowel, (b) DFT spectrum of a /t a/ token during the burst, (c) formant transition measurements, and (d) determination of formant transition offset (with the F1 frequencies obtained from an 8 kHz waveform using LPC analyses).

TABLE II. Acoustic measurements.

Name	Description	Name	Description
vbD	Voicebar duration	F3bA	F3 onset amplitude
bstD	Burst duration	F1eA	F1 amplitude at the end of transition
bstpA	Burst peak amplitude	F2eA	F2 amplitude at the end of transition
votD	VOT duration	F3eA	F3 amplitude at the end of transition
votpA	VOT peak amplitude	F1sl	F1 slope
F1D	F1 transition duration	F2sl	F2 slope
F2D	F2 transition duration	F3sl	F3 slope
F3D	F3 transition duration	F1dA	F1 amplitude change
F1b	F1 onset frequency	F2dA	F2 amplitude change
F2b	F2 onset frequency	F3dA	F3 amplitude change
F3b	F3 onset frequency	F1df	F1 frequency change
F1e	F1 frequency at the end of transition	F2df	F2 frequency change
F2e	F2 frequency at the end of transition	F3df	F3 frequency change
F3e	F3 frequency at the end of transition	F0b	F0 onset frequency
F1bA	F1 onset amplitude	F0df	F0 frequency change
F2bA	F2 onset amplitude		

formant transition onset, offset, and steady-state points [Fig. 2(c)], the following parameters were recorded: time (relative to the beginning of the utterance and measured in ms), formant frequency (from the LPC spectrum and measured in Hz), and formant amplitude (from the LPC spectrum and measured in dB). From these measurements, formant transition duration, frequency and amplitude change, and slope were calculated for each of the first three formants. Formant transition duration was defined as the time difference between the formant transition offset and onset. Using this methodology, the duration and slope of formant transitions were not necessarily the same for the three formants. Formant frequency and amplitude changes were measured between the formant transition onset and steady state (which was a more stable reference than the offset). Formant slope was calculated from formant transition onset and offset measures, since it required temporal information. These definitions imply that the formant slope does not necessarily relate to the ratio of formant frequency change to formant transition duration. Table II lists the acoustic parameters that were measured for each token.

C. Perceptual experiments

1. Participants

Listening experiments were conducted with four paid subjects (different from those who participated in the recording session), two males and two females, all native speakers of American English aged between 18 and 36 years, who passed a hearing test and participated in a training session.

2. Procedure

Experiments took place in a sound attenuating room. Digital speech stimuli were played via an Ariel Pro Port 656 board digital-to-analog converter (16 bits at a rate of 16 kHz). The resulting analog waveforms were amplified by a Sony 59ES DAT recorder. The amplified analog waveforms were then presented binaurally with identical waveforms to each ear via Telephonics TDH49P headphones. The system was calibrated within 0.5 dB (from 125 to 7500 Hz at third octave intervals) using a 6-cc coupler and a Larson Davis 800B sound level meter prior to each experiment. The sound level meter was set to the "A" weighting scale with a slow response. Pre-amp levels and digital internal level offsets were set to establish a relation between the digital RMS energy and the actual SPL level.

Each signal was played at 60 dB SPL, and the noise level was adjusted. The SPL of the speech signals were set based on its peak RMS energy on a 30 ms rectangular window. The SPL of the white Gaussian noise was adjusted based on its RMS energy to result in different SNRs. The speech signal was added to a 1250 ms noise (or silence) segment such that it was centered in the middle of the segment with equal durations of noise (or silence) before and after the speech signal. Sessions lasted for no longer than 2 hours, and subjects were instructed to take at least one break every hour. No feedback was given at any time.

The experiments were of the two alternate forced choice (2AFC) type. To counterbalance the effects of talker and token order, testing was administered in blocks of 64 pseudorandomized items (32 tokens $\times$ 2 presentations) for each CV pair (for example, /b a , p a/). When an utterance was played, participants were asked to label the sound heard as either the voiced or voiceless consonant (e.g., /b/ or /p/). A computer program was developed to record participants' responses from their keyboard inputs. The test was then repeated at different SNR levels. The order of SNR conditions was −15 dB, −10 dB, −5 dB, 0 dB, 5 dB, 10 dB, and quiet (same order for all listeners). The CV pairs were presented in the order of /ba,pa/, /bi,pi/, /bu,pu/, /da,ta/,/di,ti/, /du,tu/, /ga,ka/, /gi,ki/, and /gu,ku/.

III. VOICING CLASSIFICATION BASED ON ACOUSTIC MEASUREMENTS

The acoustic measurements were analyzed using logistic regression (Benki, 2001; Menard, 1995; Nearey, 1997), where the quiet speech tokens were classified as either voiced or voiceless according to a single acoustic property

TABLE III. Percent correct classification (shown as a superscript) of the quiet speech tokens (from all talkers) based on a single acoustic property measured without the addition of the white Gaussian noise.

/ba,pa/	/bi,pi/	/bu,pu/	/da,ta/	/di,ti/	/du,tu/	/ga,ka/	/gi,ki/	/gu,ku/
votD^{100}	votD^{100}	votD^{100}	votD^{100}	votD^{100}	votD^{100}	votD^{100}	votD^{100}	votD^{100}
F1b^{100}	F3b^{97a}		F1b^{100}	votpA^{100}	votpA^{94a}	F1b^{100}	F0b^{97a}	
F1df^{100}	F3eA^{97a}		F1df^{100}	F2b^{94a}		F2b^{100}	F0df^{94a}	
F1sl^{97}	F3sl^{94a}		F1D^{94a}	F1D^{91a}		F1df^{100}		
	F2dA^{94a}		F2b^{94a}	F3df^{91a}		F2df^{100}		
	F3df^{94a}		F2df^{94a}			F1sl^{94}		
	F2b^{91a}		F3b^{91a}			F1eA^{91a}		
	F2sl^{91}							
	F2df^{91a}							

[a]Male and female token difference was significant at a $p<0.000\,179$ level (Bonferroni corrected $p<0.050$ for multiple comparisons; male and female tokens had separate thresholds on the acoustic property for voicing classification). When male and female token difference was not significant, one threshold was used.

measured without the addition of the white Gaussian noise. A separate logistic regression model was applied to each acoustic variable for each CV pair,

$$\log[\text{prob}/(1-\text{prob})] = \alpha + \beta\,\text{Mea} + e, \quad (1)$$

where prob is the probability of a token being voiceless, α is a constant, β is a weighting coefficient, Mea is one acoustic feature, and e is the error term. Table III lists the results in terms of percent correct classification based on logistic regression using the tokens from all talkers. If the gender effect was significant (Bonferroni corrected $p<0.050$), a dummy variable representing gender difference was added in the logistic regression for a better fitting:

$$\log[\text{prob}/(1-\text{prob})] = \alpha + \beta\,\text{Mea} + \gamma G + e, \quad (2)$$

where γ is a weighting coefficient, and G represents gender (1 for male and 0 for female). The significance of gender effect was determined by

$$LR[2] = (-2LL_{\text{MF}}) - (-2LL_{\text{M}}) - (-2LL_{\text{F}}), \quad (3)$$

where the likelihood ratio LR is a distributed chi-square with 2 degrees of freedom, and LL_{MF}, LL_{M}, and LL_{F} are the log likelihood functions derived from Eq. (1) using data pooled from all talkers, male talkers, and female talkers, respectively. For each token, the plosive was either voiced or voiceless, and thus the prob in Eqs. (1) and (2) was either 0 or 1. After logistic regression, $\alpha+\beta\,\text{Mea}=0$ for Eq. (1) or $\alpha+\beta\,\text{Mea}+\gamma G=0$ for Eq. (2) was used for classification, and results were compared against ideal classification to obtain the percent correct scores. In Table III, the acoustic measures are listed according to their percent correct scores for each CV pair. Only acoustic measures with 90% or higher correct classification are listed.

Prior studies (Liberman *et al.*, 1958; Lisker and Abramson, 1964) have shown that the VOT duration (votD) is a significant acoustic cue for voicing in plosives. In the present study, the VOT duration proved to be the single best acoustic property for classification of voicing for the quiet speech tokens (Table III). All tokens could be correctly classified based solely on the VOT duration (see Fig. 3) with voiceless plosives having a significantly longer VOT duration than voiced plosives [$t(286)=44.470$, $p=0.000$]. The VOT duration was less than 20 ms for most of the voiced tokens, and over 45 ms for most of the voiceless tokens. Thus, a VOT duration of 30 ms could be considered as a voicing boundary, except for the /gi,ki/ pair, which had a VOT duration boundary of 40 ms. Another VOT property, VOT peak amplitude (votpA), which can be considered a measurement of the loudness of aspiration, produced relatively high percentage of correct identifications for the alveolar pairs /di,ti/ and /du,tu/.

Voicing classification based solely on the duration of the voicebar (vbD) did not result in a high percentage of correct classification results for all talkers, whereas it was an important factor in voicing classification for seven out of the nine CV pairs from female talkers. While the presence of a voicebar is almost certainly indicative of a voiced utterance, the converse does not hold. For all but one token, voiceless tokens showed no voicebar. Only two of the 72 voiced tokens spoken by the male talkers showed a voicebar, while most (65 out of 72) of the voiced tokens spoken by the female talkers showed a voicebar. This, however, does not necessarily indicate that gender plays an important role in whether or not there is a voicebar. Since there were only four talkers, the results can be attributed to individual differences. Previous studies have shown that some talkers tend to produce a voicebar while others do not (Lisker and Abramson, 1964).

The loudness of the burst, including intensity and duration, has been cited as a possible cue for voicing (Lisker and Abramson, 1964; Repp, 1979). However, the present study shows that classification based solely on the burst properties did not produce high percent correct classification. The loudness of the burst (peak amplitude; bstpA) was about 2.3 dB higher on average for voiceless tokens than for their voiced counterparts, but the range of values between voiced and voiceless tokens overlapped greatly. Burst measurements were short (about 5 ms) for both voiced and voiceless labial tokens. In general, burst duration (bstD) was more related to place of articulation and vowel context than to voicing distinctions: Velars appeared to have longer burst durations than labials and alveolars (about 10 ms longer in the /a/ context and 20 ms longer in the /i/ and /u/ contexts).

Summerfield and Haggard (1977) showed that F1 onset frequency is important in the context of a low vowel (e.g., /a/), but not in high vowel contexts (e.g., /i/ and /u/); similar

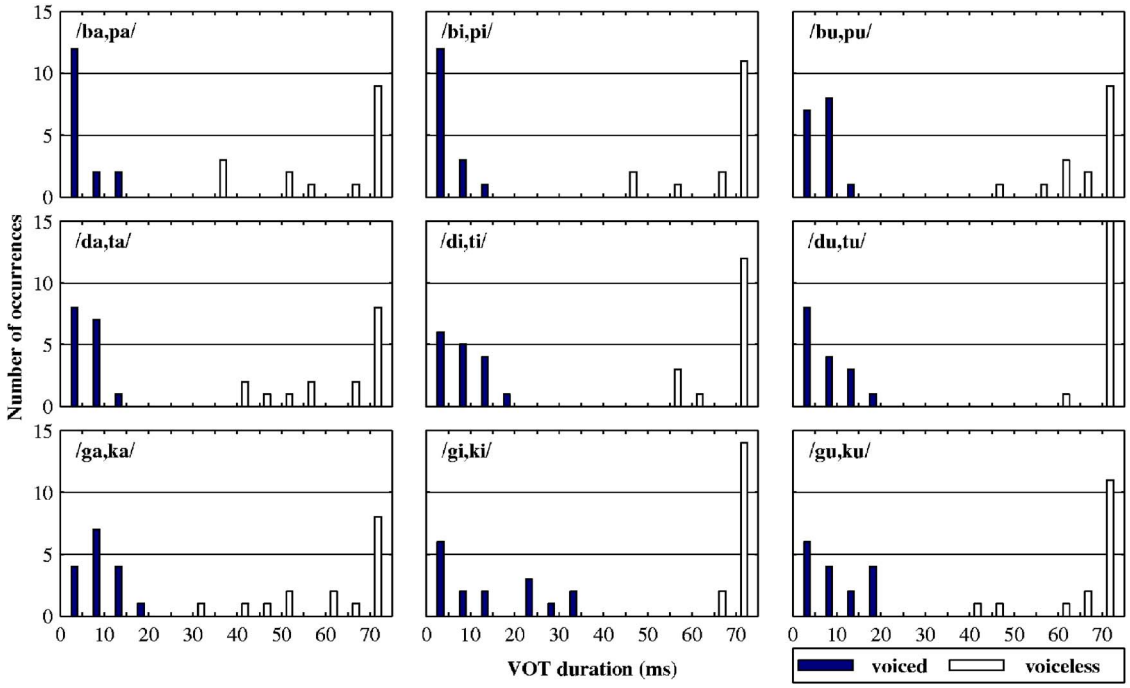


FIG. 3. Histograms of VOT duration (votD) for the nine voiced/voiceless pairs with the voiced and voiceless tokens counted separately. The histogram bin centers range from 2.5 to 72.5 ms with a 5 ms step. VOT duration of more than 75 ms is counted into the 72.5 ms center region.

results are shown in Table III. Table III shows that the F1 transition was an excellent classifier for voicing only in the /a/ context. All /Ca/ tokens could be correctly classified based solely on their F1 onset frequency (F1b, see Fig. 4). In the /a/ context, voiceless tokens usually showed an F1 onset frequency that was higher than 600 Hz, while voiced tokens were below that [mean absolute difference=379 Hz, $t(94)=15.281$, $p=0.000$]. For tokens in the /i/ and /u/ contexts, however, the F1 onset frequency was in about the same range for both voiced and voiceless tokens with mean absolute differences of 43 and 33 Hz, respectively. As mentioned earlier, the F1 transition has been considered an important property for voicing classification (Stevens and Klatt, 1974). Accordingly, the F1 frequency change (F1df) was significantly different for voiced and voiceless tokens only in the /a/ context [mean absolute difference=360 Hz, $t(94)=15.915$, p

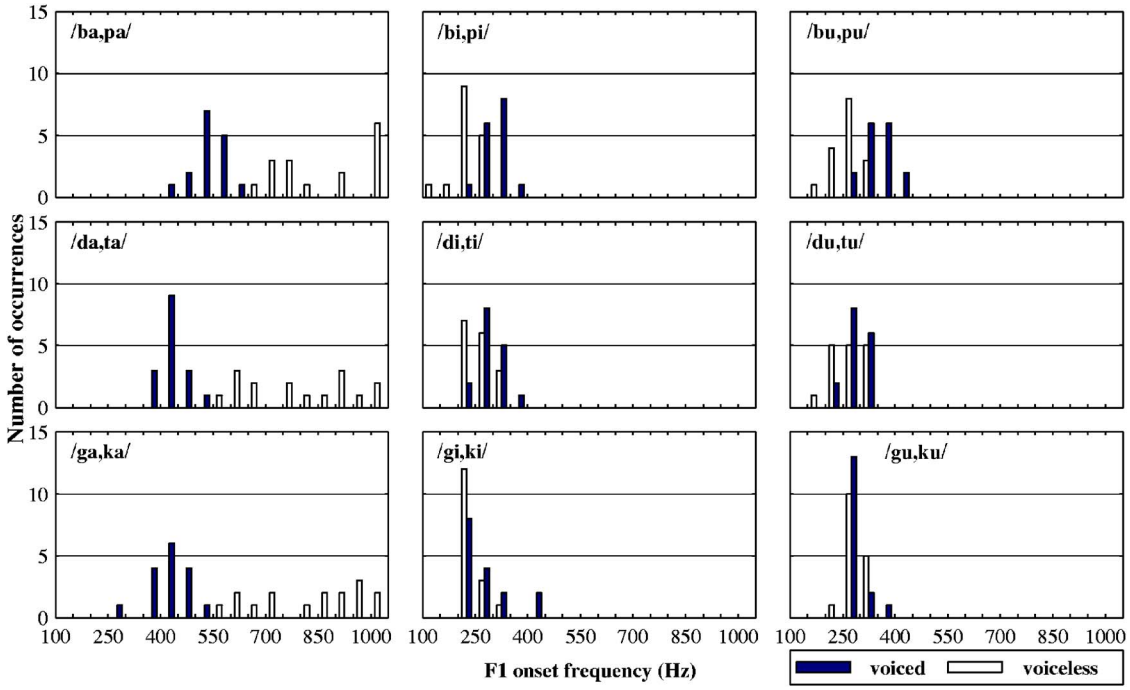


FIG. 4. Histograms of F1 onset frequency (F1b) for the nine voiced/voiceless pairs with the voiced and voiceless tokens counted separately. The histogram bin centers range from 125 to 1025 Hz with a 50 Hz step. F1b of less than 100 Hz and of more than 1050 Hz is counted into the 125 Hz center and 1025 Hz center regions, respectively. Note that the bins are well separated by voicing categories only in the /a/ context.

TABLE IV. Percent correct voicing judgments as a function of SNR, place of articulation, and vowel context (data averaged across all talkers and all listeners).

	/b,p/			/d,t/			/g,k/		
SNR (dB)	/a/	/i/	/u/	/a/	/i/	/u/	/a/	/i/	/u/
Quiet	100	100	99	100	100	99	97	100	100
10	98	93	89	98	99	98	98	95	99
5	98	90	82	100	95	89	97	93	89
0	95	82	83	97	84	86	96	91	84
−5	88	70	74	96	75	69	91	80	80
−10	66	55	66	62	61	60	57	59	69
−15	47	49	53	44	51	50	49	50	59

$=0.000$]. For tokens in the /i/ and /u/ contexts, the F1 frequency change was in about the same range for both voiced and voiceless tokens with mean absolute differences of 42 and 32 Hz, respectively. However, given the covariation between F1b and F1df and the results by Summerfield and Haggard (1977), F1 onset frequency (F1b), instead of F1 frequency change, was emphasized in this study. The percent correct classification from the F1 slope measurements (F1sl) was not as high as the onset frequency or frequency change of F1.

For voiceless plosives, a large portion of the formant transition occurs during the VOT and appears in the aspiration noise. Thus, voiceless tokens would exhibit short transition durations (in the vowel). Voiced plosives, on the other hand, are voiced throughout the transition and should therefore have longer formant transition durations. In this study, however, F1 transition duration (F1D) was not an accurate classifier for voicing, except for /da,ta/ and /di,ti/. F1 amplitude at the end of transition (F1eA) produced better voicing classification for the /a/ context than for the /i/ and /u/ contexts, but only the /ga,ka/ pair received above 90% correct classification based on F1eA. The F1 amplitude change measurements (F1dA) resulted in poor classification. F2 and F3 measurements were generally poor classifiers for voicing, while notable exceptions were F2 and F3 measurements for /da,ta/, /ga,ka/, /bi,pi/, and /di,ti/ (Table III).

On average, F0 onset frequency (F0b) was about 20 Hz higher for voiceless tokens than for voiced ones; however, the overlap between voiced and voiceless tokens was considerable. Similarly, F0 frequency change (F0df, i.e., an F0 drop) tended to be slightly greater (15.7 Hz) for voiceless tokens than for voiced ones but with a large overlap. Voicing classification based solely on F0b and F0df produced a high percentage of correct results only for /gi,ki/ (97% and 94%, respectively) with a significant gender effect (Bonferroni corrected $p<0.050$).

Note that in Table III, when male and female token differences were significant, separate thresholds were used. This was achieved by adding gender as a variable in logistic regression [see Eq. (2)]. Therefore, high percent correct classification can still be achieved even when there is a significant male and female token difference.

In summary, the VOT duration was the single-best classifier in voicing for all nine CV pairs. The properties of the first formant frequency (F1 onset frequency, F1 frequency change, and F1 slope) were good classifiers only for the three /Ca/ pairs. Because the /Ca/ pairs exhibited stronger F1 cues than the /Ci/ and /Cu/ pairs, we expect them to have better voicing perception in noise than the /Ci/ and /Cu/ pairs. F1 transition duration was a good classifier only for /da,ta/ and /di,ti/. Properties of F2 and F3 were good classifiers for four CV pairs. Voicebar duration was a reliable classifier for seven out of the nine CV pairs from the female talkers.

IV. PERCEPTUAL RESULTS

A. Percent correct classification

The percentage of correct voicing judgments was computed and listed as a function of SNR, place of articulation, and vowel context. The percent correct values shown in Table IV were calculated using all the data collected from the perceptual experiments, including all listeners and all talkers. Each data entry thus represents 256 responses (four talkers × four listeners × four recordings of each CV × two presentations × two consonants).

The listeners appeared to have had a particularly difficult time classifying the /bu,pu/ pair (with 89% correct voicing judgments) even when the SNR was 10 dB. However, for CV pairs other than /bu,pu/, the percent correct of voicing judgments was 93% or above when the SNR was 10 dB. Among the nine CV pairs, /da,ta/ yielded the best voicing judgment performance (96% correct) when the SNR was −5 dB. For SNRs of −10 dB and below, voicing judgments for all nine pairs were dramatically affected by noise (below 70% correct).

A four-way repeated measures analysis of variance (ANOVA) (gender × place of articulation × vowel × SNR) was used to analyze the perceptual results after an arcsine transformation was applied. The effects of vowel context [$F(2,6)=35.254$, $p=0.000$], SNR [$F(6,18)=341.273$, $p=0.000$], place of articulation [$F(2,6)=14.205$, $p=0.005$], and gender of the talker [$F(1,3)=13.066$, $p=0.036$] were significant. The mean percent correct for vowel context, SNR, and place of articulation is plotted in Fig. 5. As expected, the vowel /a/ context yielded higher percent correct classification than the /i/ context [$F(1,3)=35.573$, $p=0.009$], but there was no significant difference between the /i/ and /u/ contexts [$F(1,3)=1.473$, $p=0.312$; see Fig. 5(a)]. Also as expected, the number of correct responses decreased as the SNR level decreased ($p<0.050$), and decreased more quickly after the SNR was reduced below −5 dB. Figure 5(b) shows that most of the nine CV pairs had 100% or close to

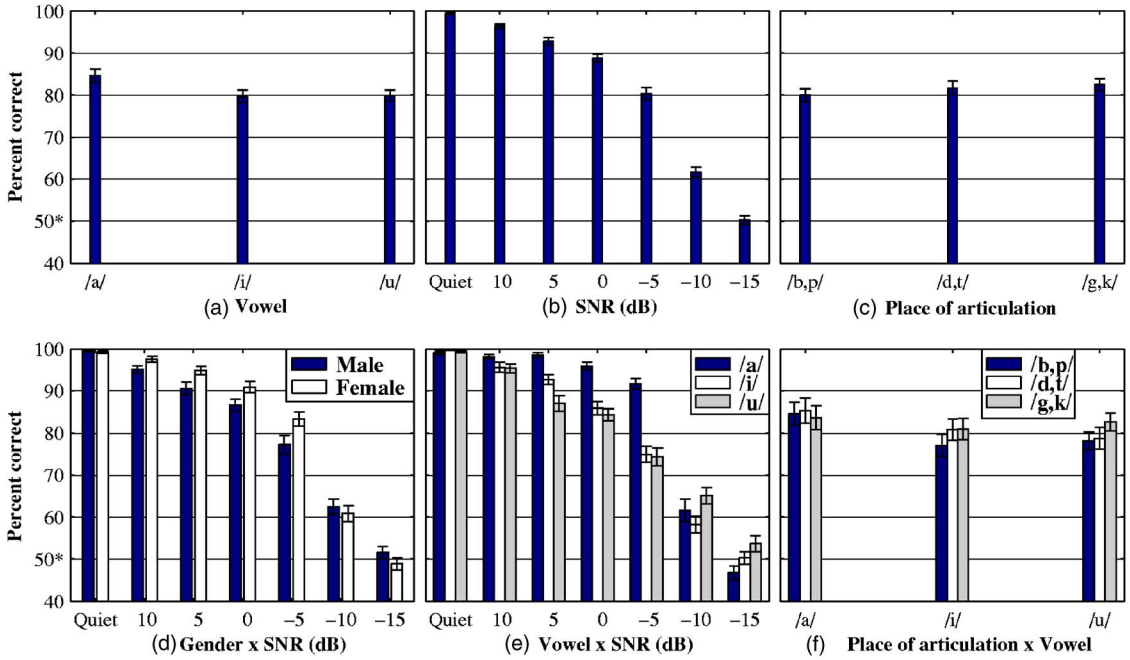


FIG. 5. Percent correct voicing judgments as a function of (a) vowel context (/a/, /i/, or /u/), (b) SNR (dB), (c) place of articulation (/b,p/, /d,t/, or /g,k/), (d) gender (male or female) and SNR combination, (e) vowel and SNR combination, and (f) place of articulation and vowel combination. The error bars show standard errors of means. The 50% chance performance is indicated with 50* on the percent correct axis.

100% correct voicing judgments in the absence of noise. Voicing discrimination degraded significantly between 0 and −10 dB SNRs. When the SNR was −15 dB, the percent correct of voicing judgments was about 50%, which is chance performance. This is consistent with the Miller and Nicely (1955) study that indicated voicing was still discriminable at an SNR of −12 dB.

As for place of articulation [Fig. 5(c)], percent correct scores for alveolars were significantly higher than those of bilabials [$F(1,3)=14.706$, $p=0.031$] but not significantly different from those of velars [$F(1,3)=0.002$, $p=0.964$].

The interactions between gender and SNR [$F(6,18)=6.452$, $p=0.001$], between vowel and SNR [$F(12,36)=16.867$, $p=0.000$], between place of articulation and SNR [$F(12,36)=2.945$, $p=0.006$], and between place of articulation and vowel [$F(4,12)=4.067$, $p=0.026$] were significant, while the interactions between gender and place of articulation [$F(2,6)=0.044$, $p=0.957$] and between gender and vowel [$F(2,6)=0.127$, $p=0.883$] were not significant. The gender and SNR interaction, vowel and SNR interaction, and place of articulation and vowel interaction effects are shown in Fig. 5. The interactions of more than two factors (gender, place of articulation, vowel, and SNR) were not significant ($p>0.050$) except for place of articulation × vowel × SNR interaction [$F(24,72)=2.027$, $p=0.011$]. Figure 5(d) shows an interesting gender and SNR interaction effect. Voicing judgments for stimuli from male talkers degraded steadily with decreasing SNR, whereas for female talkers, the performance of voicing discrimination gradually degraded with decreasing SNR when the SNR was above −5 dB and then degraded dramatically for lower SNRs. A possible reason for the gender × SNR interaction may be due to the differences in F1 onset frequency between voiced and voiceless CVs for male and female talkers. Generally speaking, female talkers have shorter vocal tracts than male talkers, and thus produce higher formant frequencies including F1 onset frequencies for voiceless CVs, which were shown to influence voicing judgments (Summerfield and Haggard, 1977). For example, female talkers produced higher F1 onset frequencies in voiceless /Ca/ and /Ci/ tokens than male talkers.

Figure 5(e) shows that the /a/ context was the most robust one, most likely because the F1 differences between the voiced and voiceless plosives are prominent only in the /a/ context. These differences may have helped listeners perceive voicing more robustly. The figure also shows that voicing judgments for stimuli in the /i/ and /u/ contexts degraded steadily with decreasing SNR, whereas in the /a/ context, the performance of voicing discrimination gradually degraded with decreasing SNR when the SNR was above −5 dB and then dropped rapidly for lower SNRs.

Figure 5(f) shows the vowel and place of articulation interaction effect. For example, voicing classification was better for velars than for bilabials and alveolars in the /i/ and /u/ contexts.

In summary, the vowel /a/ context yielded higher percent correct voicing classification than the /i/ and /u/ contexts. Velars yielded higher percent correct classification than bilabials and alveolars in the /i/ and /u/ contexts. Voicing judgments for stimuli from male talkers (or the /i/ and /u/ contexts) degraded steadily with decreasing SNR, whereas for female talkers (or the /a/ context), voicing discrimination degraded dramatically when the SNR was below −5 dB.

B. Threshold SNRs for voicing classification in noise

A traditional approach for investigating voicing distinction has been to explore the perceptual boundaries between

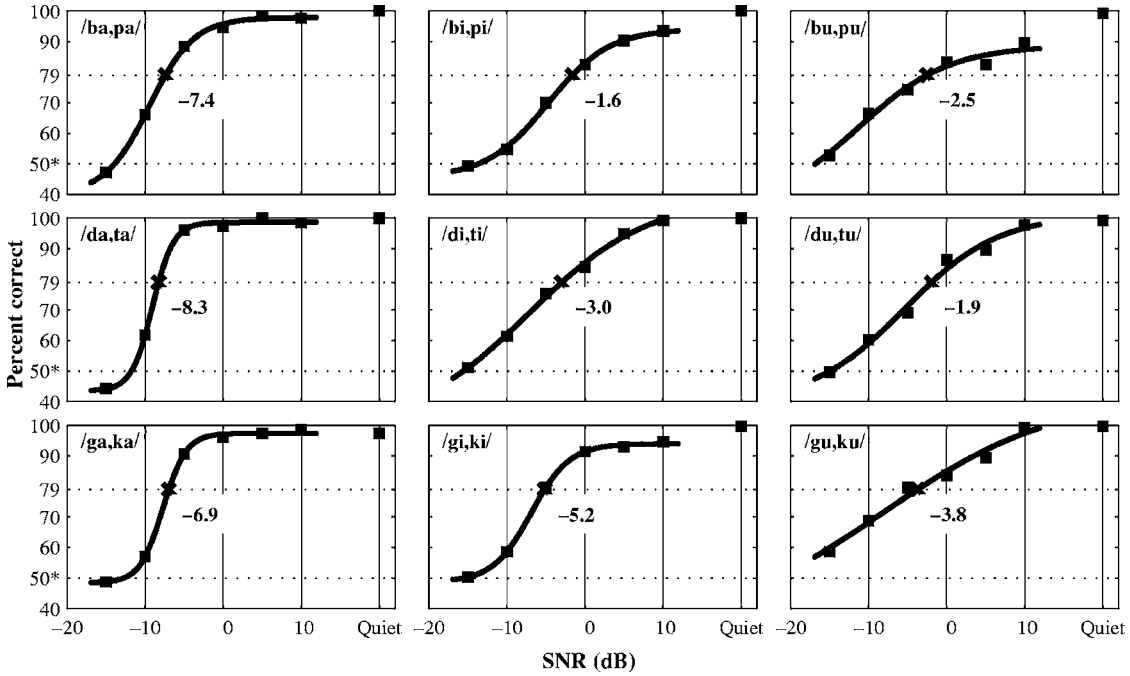


FIG. 6. A sigmoid fitting of percent correct scores as a function of SNR (dB) for the nine voiced/voiceless pairs. For each voiced/voiceless pair, the 79% threshold line is drawn, and the threshold SNR value is labeled. The 50% chance performance is indicated with the dash lines and with a 50* on the percent correct axis.

phonetic categories (e.g., VOT, F1, VOT × F1, etc.). In order to analyze how the acoustic properties account for the perceptual results, a single value for each CV pair was needed to represent the robustness of that CV pair in the presence of noise. That value, or threshold, was computed along the SNR continuum. The data for the nine CV pairs were arranged into plots as shown in Fig. 6 where percent correct is plotted versus SNR. A sigmoid was then fit to each plot (excluding data for the quiet condition) and described by the following equation:

$$y = c + \frac{d-c}{2}\left(1 - \frac{1-e^{(x-b)/a}}{1+e^{(x-b)/a}}\right), \tag{4}$$

where x represents SNR and y represents percent correct. The parameters a (rate of change of the sigmoid), b (halfway point), c (bottom of the sigmoid), and d (top of the sigmoid) were varied systematically to obtain the best fit sigmoid by minimizing the mean squared error. Note that in a regular logistic regression model, c and d are fixed to 0 and 1, respectively, and the model fitting is based on the maximum likelihood principle. Theoretically, the percent correct versus SNR curve should be flat at about 100% for very high SNR levels, flat at about 50% (chance performance) for very low SNR levels, and monotonically increasing in between. The characteristics of a sigmoid match these requirements, and it was therefore chosen as the curve that best represented the data. From the best fit sigmoid, the threshold SNR level corresponding to 79% correct responses was obtained (Chen, 2001; Hant, 2000; Levitt, 1971). Thus, a single threshold SNR value for each of the nine pairs of voiced/voiceless CVs was calculated to represent the perceptual robustness of that pair. A lower threshold SNR corresponded to better perceptual results (more robust to noise).

Threshold SNR levels for all CV pairs corresponding to 79% correct responses averaged over all talkers are shown in Fig. 6. Threshold SNRs computed using data separated by gender of the talker are shown in Fig. 7. Figure 6 shows that for all talkers, threshold SNRs for CVs in the /i/ and /u/ contexts were lower (more robust) for velars (ranging from −5.2 to −3.8 dB) than for labials and alveolars (ranging from −3.0 to −1.6 dB). CVs in the /a/ context (Group 1 in Fig. 7) appeared to be significantly more robust in noise than those in the /i/ [$t(10)=4.846$, $p=0.001$] and /u/ [$t(10)=5.556$, $p=0.000$] contexts. This effect agrees with the results shown in Fig. 5(e). In most cases, the tokens from female talkers were more perceptually robust than those from male talkers [paired $t(8)=3.475$, $p=0.008$; much lower threshold SNRs (by over 3 dB) for /di,ti/, /bu,pu/, and /du,tu/ pairs; labeled as Group 2 in Fig. 7]. Such effect is consistent with that shown in Fig. 5(d).

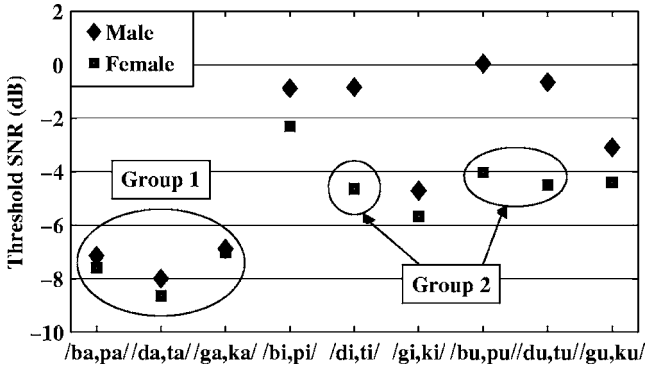


FIG. 7. Threshold SNR values (dB) for the nine voiced/voiceless pairs, separated by gender of the talkers. Group 1 (/Ca/ syllables) shows little difference in perceiving male versus female tokens, while Group 2 (/di,ti/, /du,tu/, and /bu,pu/) shows large gender related differences.

Voicing classification for /gi,ki/ was more robust than those for /bi,pi/ and /di,ti/ (Fig. 6). Threshold SNRs for /gu,ku/ were lower than those for /bu,pu/ and /du,tu/ when the talker was a male, while the values were very close when the talker was a female (Fig. 7).

In summary, CVs in the /a/ context appeared to be more robust in noise than those in the /i/ and /u/ contexts. Threshold SNRs for CVs in the /a/ context were lower (more robust) for alveolars than for labials and velars. Threshold SNRs for CVs in the /i/ and /u/ contexts were lower for velars than for labials and alveolars. In most cases, the tokens from female talkers were more noise robust than those from male talkers. This can be attributed in part to differences in formant frequencies and F0 frequencies in addition to the existence of a voicebar for 65 out of the 72 voiced CV tokens from the female talkers.

V. CORRELATIONS BETWEEN THRESHOLD SNR VALUES AND ABSOLUTE ACOUSTIC DIFFERENCES OF THE MEANS

Correlations were computed between the nine threshold SNR values from the perceptual experiments and the absolute differences of the mean values of a measured acoustic property for the nine voiced/voiceless pairs (three places of articulation × three vowel contexts). The mean value of each acoustic measurement for every CV syllable was calculated from 16 tokens (four talkers × four tokens of the same syllable). The correlation is defined as

$$r = \mathrm{corr}(|\overline{\mathrm{Mea}}_v - \overline{\mathrm{Mea}}_{vl}|, 10 - \mathrm{SNR}_t) = -\mathrm{corr}(|\overline{\mathrm{Mea}}_v - \overline{\mathrm{Mea}}_{vl}|, \mathrm{SNR}_t), \quad (5)$$

where corr represents the Pearson correlation function, v represents voiced tokens, vl represents voiceless tokens, Mea represents one type of acoustic measurement, the bar over Mea represents the mean operation, SNR_t represents the threshold SNR values, and $10-\mathrm{SNR}_t$ indicates how much the threshold SNRs were below 10 dB. The absolute difference of the means was chosen under the assumption that the greater the distance between the means, the larger the separation between the associated distributions, and thus the more distinct the acoustic property in question is for that voiced/voiceless CV pair. Thus, if an acoustic property is an important cue for voicing, then a larger absolute difference between the means would correspond to better performance (a lower threshold SNR), while a smaller absolute difference between the means would correspond to poorer performance (a higher threshold SNR). Correlation using the absolute difference of the means has several shortcomings, but it provides a simple method for obtaining a numerical measure of how well the perceptual results correlate with acoustic properties. Pearson product correlation coefficients were obtained only for those acoustic properties that appear in Table III.

Table V lists the results of correlating threshold SNRs (using data from all talkers) with the absolute differences of the means of several acoustic properties. The first formant measurements yielded the highest and most significant correlations, F1 onset frequency (F1b) and F1 transition frequency change (F1df) having correlation coefficients of 0.86 (Bonferroni corrected $p=0.048$) and 0.87 (Bonferroni corrected $p=0.044$), respectively. A close examination of F1 measurements indicated that the absolute differences of means of F1b and F1df were perfectly correlated [$r(7)=0.999$, $p=0.000$]. Given such covariation between F1b and F1df, we may interpret the results of Summerfield and Haggard (1977) to mean that the F1 onset frequency, rather than the F1 frequency change, was the most important cue for voicing perception at low SNRs. F1 amplitude at the end of transition (F1eA) also yielded a relatively high correlation coefficient of 0.84 (Bonferroni corrected $p=0.085$). Summerfield and Haggard (1977) showed that the F1 amplitude was not an important cue for voicing in quiet conditions. In this study, in quiet, F1eA signaled voicing prominently only for /g a , k a/ (91% correct; Table III). In noise, however, F1eA might become important in voicing perception as demonstrated by a high correlation coefficient (Table V). F1 slope (F1sl) and F1 transition duration (F1D) also showed relatively high correlation coefficients of 0.79 and 0.78, respectively. These correlations occurred because the /a/ sound is the most robust in noise, and because simultaneously the /a/ context has the largest absolute F1 onset frequency difference for the voiced and voiceless syllable-initial plosives.

TABLE V. Correlation coefficients (shown as superscripts) of threshold SNRs with absolute differences of the means of various acoustic properties across all talkers (with the highest correlation coefficient listed first).

$\mathrm{F1df}^{0.87}$[a]	$\mathrm{F2df}^{0.34}$	$\mathrm{F0b}^{-0.29}$
$\mathrm{F1b}^{0.86}$[a]	$\mathrm{F3df}^{0.08}$	$\mathrm{F2s1}^{-0.40}$
$\mathrm{F1eA}^{0.84}$	$\mathrm{F3b}^{0.01}$	$\mathrm{F2dA}^{-0.45}$
$\mathrm{F1s1}^{0.79}$	$\mathrm{F3s1}^{-0.06}$	$\mathrm{F3eA}^{-0.50}$
$\mathrm{F1D}^{0.78}$	$\mathrm{votpA}^{-0.08}$	$\mathrm{votD}^{-0.85}$
$\mathrm{F2b}^{0.46}$	$\mathrm{F0df}^{-0.14}$	

[a]The correlation was significant at a $p<0.002\,94$ level (df=7, Bonferroni corrected $p<0.050$ for multiple comparisons).

VOT duration (votD), on the other hand, showed a highly negative correlation with the perceptual results [$r(7)=-0.85$, Bonferroni corrected $p=0.060$]. A negative correlation coefficient indicates a larger distance between the means correlated with a higher (worse) threshold SNR. The VOT duration cue is easily corrupted by noise since the burst and aspiration parts are of low amplitude compared to the vowel onset.

F0 measurements (F0b and F0df) yielded negative but small-amplitude correlation coefficients. Thus, F0 cues might be easily disrupted in the voicing perception in noise.

Some of the correlations were further examined as shown in Fig. 8; the panels show threshold SNRs versus the absolute differences of the means for several acoustic properties. Numbers inside the panels are the correlation coefficients between the SNRs and the absolute differences of the means. For example, the absolute differences of the means of F1 onset frequency (F1b) and F1 amplitude at the end of transition (F1eA) were large for the /a/ context and small for the /i/ and /u/ contexts. Thus, the high correlation for F1 onset frequency indicated that the threshold SNRs were much lower for /Ca/'s than for /Ci/'s and /Cu/'s, which is

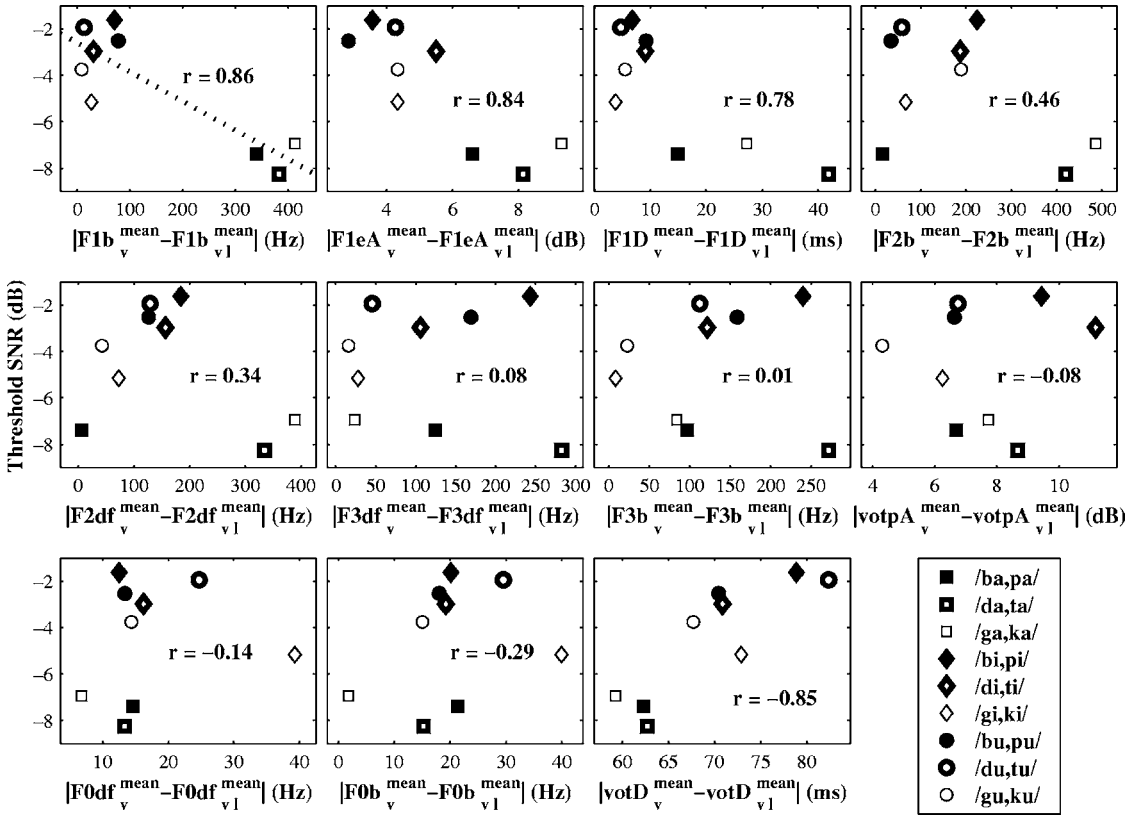


FIG. 8. Threshold SNRs and absolute differences of the means of several acoustic properties for the nine voiced/voiceless pairs. For each point (CV pair), the threshold SNR and the difference of the means resulted from the perception and measurement of 32 tokens (four talkers × four tokens of the same syllable × two syllables in one CV pair), respectively. Each point is indicated with its vowel context [/a/ (square), /i/ (diamond), or /u/ (circle)] and place of articulation [/b,p/ (solid), /d,t/ (bold), or /g,k/ (thin)]. Numbers inside panels are the correlation coefficients between the nine threshold SNRs and the nine absolute differences of the means. A fitted line is shown in only one panel for illustrative purposes.

consistent with Fig. 6. As shown in Fig. 8, the highly negative correlation of threshold SNRs with VOT duration measures occurred because the CV pairs that were best separated by the VOT duration (all /d,t/ and /b,p/ pairs) also exhibited the highest (worst) threshold SNRs. In the speech production process under quiet recording conditions, there is a trading relation between VOT duration and F1 onset frequency (Summerfield and Haggard, 1977). That is, CV pairs with larger absolute VOT duration difference are produced with smaller F1 onset frequency differences, and vice versa (as is evident in Fig. 8). Because the VOT duration is masked at a higher SNR than the F1 onset frequency, the CV pairs with large VOT duration differences (and correspondingly small F1 onset frequency differences) are mistakenly perceived at a higher SNR than the CV pairs with small VOT duration differences (and correspondingly large F1 onset frequency differences).

In summary, the F1 onset frequency measurements yielded a significant positive correlation with voicing perception in noise. Because the /a/ context resulted in larger absolute differences in the F1 onset frequency measurements between the voiced and voiceless plosives than the /i/ and /u/ contexts, the voicing judgments in the /a/ context were more robust than those in the /i/ and /u/ contexts. VOT duration, on the other hand, showed a highly negative correlation.

VI. GENERAL DISCUSSION

The present study examines the acoustic correlates and perception in noise of voicing in naturally spoken syllable-initial plosives. It is important to determine how human listeners trade off various acoustic cues (e.g., VOT and F1 onset frequency) at different SNR levels. The results demonstrate that the perception of voicing in plosives in noise is affected by an interaction between SNR on the one hand, and factors that affect the acoustic characteristics of the plosive release on the other hand; such factors include talkers' gender, place of articulation, and vowel context. Specifically, post-hoc correlation analyses suggest that VOT duration is the cue most predictive of voicing decisions in quiet conditions, but VOT duration appears to be masked at low SNRs, with a contrasting result that F1 onset frequency is a better voicing cue at low SNRs.

Specifically, in quiet conditions, all of the nine CV pairs were correctly classified at or near 100%. The VOT duration proved to be the single best acoustic property for voicing classification in syllable-initial plosives, and all tokens could be correctly classified based on their VOT durations. Furthermore, VOT duration boundaries differed across places of articulation, which agrees with results reported by Lisker (1975). VOT peak amplitude produced relatively high percent correct classification for two CV pairs. The F1 onset frequency was an excellent classifier for voicing only in the

/a/ context. All /Ca/ tokens could be correctly classified based solely on their F1 onset frequencies or F1 frequency changes that covaried with F1 onset frequencies. Tokens without a prominent F1 onset frequency cue can be correctly identified, indicating that the F1 onset frequency is not a necessary cue for voicing, at least not in the /i/ and /u/ contexts. Obviously, there are multiple and redundant cues for voicing. In the presence of noise, the relative roles of different acoustic cues appear to change as a function of SNR.

For noisy speech, as expected, the voicing distinction was more difficult to perceive as the SNR level decreased. Listeners could still make correct voicing judgments even when the SNR level was −10 dB. However, for an SNR of −15 dB, listeners' responses were equivalent to random guesses (chance performance). These results are consistent with the Miller and Nicely (1955) study, which concluded that voicing is robust to noise, while place of articulation is not.

For voicing classification in noise, vowel effect was significant in the sense that voicing judgments were always more accurate in the /a/ context than in the /i/ and /u/ contexts (threshold SNRs were lower for /Ca/'s than for /Ci/'s or /Cu/'s). In addition, voicing classification in /Ca/ syllables degraded gradually with decreasing SNR and then degraded rapidly for SNRs lower than −5 dB. In contrast, voicing classification in the /i/ and /u/ contexts degraded steadily across different SNR levels. Correlation analyses showed that the VOT duration contributed negatively to these threshold SNR differences. Instead, the F1 onset frequency differences between voiced and voiceless CVs for the nine pairs were highly correlated with threshold SNRs from the perceptual experiments. The highly negative correlation for the VOT duration and the significant positive correlation for the F1 onset frequency agree with the trading relationship between VOT duration and F1 onset frequency reported in Summerfield and Haggard (1977). The positive correlation for F1 onset frequency indicates that the F1 onset frequency is more important for the perception of voicing at low SNRs than the VOT duration. The high positive correlation occurred because the differences in F1 onset frequency between voiced and voiceless tokens were large and discriminative in the /a/ context, but not as large or discriminative in the /i/ or /u/ context (a perfect voiced/voiceless indicator only for CV syllables in the /a/ context). This is understandable because the information for the plosive consonants is not limited to a single time instant (Liberman *et al.*, 1967), but extends to the following vowel through coarticulation. The vowel /a/ has a high F1 onset frequency for a voiceless token and a low F1 onset frequency for a voiced token, and thus the difference in F1 onset frequency is prominent. The range of F1 onset frequency (or amplitude) change is much larger in the /a/ context than in the /i/ and /u/ contexts. For plosives, the release burst and aspiration consist of noise, which is weaker in amplitude than the vowel formants and is easily corrupted by noise, especially broadband noise. Therefore, in noisy speech, F1 onset frequency is a more dominant cue for voicing than VOT, and voicing perception is dependent on vowel context.

Similarly there was a significant gender and SNR interaction effect. Threshold SNR values for /di,ti/, /bu,pu/, and /du,tu/ with female talkers were significantly lower than those for their male counterparts' tokens (by over 3 dB). Such an interaction effect could be attributed to the fact that the properties of fundamental frequency and formant frequencies (or transitions) for female talkers are different from those for male talkers. For example, the absolute differences of means of F1 onset frequency in the /Ca/ voiced/voiceless pairs were larger for female talkers than for male talkers. In addition, a voicebar was present in the majority of voiced tokens by the female talkers, while only two out of 72 voiced tokens by the male talkers showed a voicebar. Further, for female talkers, distinct differences in F1 transition duration between the voiced and voiceless tokens were present in /di,ti/, /bu,pu/, and /du,tu/, but not for /gu,ku/. Some of the /Ci/'s and /Cu/'s from the male talkers were well classified using the F1 transition duration measurements, but they did not have low perceptual threshold SNRs. This may be due to their short F1 transitions. A longer F1 transition is more easily detectable, particularly in noise (Hant, 2000). This is, however, only useful if a voicing cue is present in the F1 transition. Therefore, for noisy speech, better voicing classification results were obtained only if the F1 transition contained distinct differences between the voiced and voiceless CVs, and if the F1 transition duration was relatively long (over 10 ms). Furthermore, for /di,ti/, /bu,pu/, /du,tu/, and /gu,ku/, the F1 transition duration was much longer for tokens generated by the female talkers (6 to 20 ms) than by the male talkers (1 to 10 ms), for both the voiced and voiceless CVs. This could be another explanation for the female tokens being more noise robust.

Results in this study also indicate that F0 differences between voiced and voiceless plosives are not critical cues for voicing perception in additive white Gaussian noise.

Threshold SNRs for CVs in the /i/ and /u/ contexts were lower (more robust) for velars (/g,k/) than for labials (/b,p/) and alveolars (/d,t/). One possible explanation for the better performance of the velars (in the absence of an F1 transition cue) is that the differences of F0 frequency change, F0 onset frequency, and F1 amplitude at the end of transition between voiced and voiceless tokens, although not prominent, were larger for /gi,ki/ and /gu,ku/ than for /bi,pi/, /di,ti/, /bu,pu/, and /du,tu/.

In future studies, experiments will be conducted using a larger data set, with more talkers, in order to determine whether differences in gender that appeared in this study were in fact due to gender, or due to individual differences. Perceptual experiments could also be conducted using synthetic stimuli to construct acoustic continua and control interactions between the various acoustic properties (e.g., independently vary VOT, F1 onset frequency, F1 frequency change, and F1 amplitude). Given that the F1 onset frequency is an important cue for voicing perception at low SNRs, an interesting future research direction would be to investigate whether talkers will enhance the F1 onset frequency difference under noisy recording conditions to emphasize the voicing contrast.

ACKNOWLEDGMENTS

This work was supported in part by an NSF Post-Graduate Fellowship to Marcia Chen and by an NIH grant. The authors also thank Ann Bradlow and three anonymous reviewers for constructive comments and suggestions. The views expressed here are those of the authors and do not necessarily represent those of the National Science Foundation and the National Institute of Health.

Allen, J. B. (**1994**). "How do humans process and recognize speech?" IEEE Trans. Speech Audio Process. **2**, 567–577.

Benki, J. (**2001**). "Place of articulation and first formant transition pattern both affect perception of voicing in English," J. Phonetics **29**, 1–22.

Chen, M. (**2001**). "Perception of voicing for syllable-initial plosives in noise," Master thesis, Electrical Engineering Department, University of California at Los Angeles.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: evidence from 18 languages," J. Phonetics **27**, 207–229.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Fitch, H. L., Halwes, T., Erickson, D. M., and Liberman, A. M. (**1980**). "Perceptual equivalence of two acoustic cues for stop-consonant manner," Percept. Psychophys. **27**, 343–350.

Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.

Hall, M. D., Davis, K., and Kuhl, P. K. (**1995**). "Interactions between acoustic dimensions contributing to the perception of voicing," J. Acoust. Soc. Am. **97**, 3416.

Hant, J. (**2000**). "A computational model to predict human perception of speech in noise," Ph.D. dissertation, Electrical Engineering Department, University of California at Los Angeles.

Hant, J., and Alwan, A. (**2000**). "Predicting the perceptual confusion of synthetic plosive consonants in noise," Proceedings of the Sixth International Conference on Spoken Language Processing, Beijing, China, pp. 941–944.

Hant, J., and Alwan, A. (**2003**). "A psychoacoustic-masking model to predict the perception of speech-like stimuli in noise," Speech Commun. **40**, 291–313.

Hermansky, H. (**1990**). "Perceptual linear prediction (PLP) analysis for speech," J. Acoust. Soc. Am. **87**, 1738–1752.

Kewley-Port, D. (**1982**). "Measurement of formant transitions in naturally produced stop consonant-vowel syllables," J. Acoust. Soc. Am. **72**, 379–389.

Klatt, D. H. (**1975**). "Voice onset time, friction, and aspiration in word-initial consonant clusters," J. Speech Hear. Res. **18**, 686–706.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.

Liberman, A. M., Delattre, P. C., and Cooper, F. S. (**1958**). "Some cues for the distinction between voiced and voiceless stops in initial position," Lang Speech **1**, 153–167.

Lisker, L. (**1975**). "Is it VOT or a first-formant transition detector?" J. Acoust. Soc. Am. **57**, 1547–1551.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lisker, L., and Abramson, A. S. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," *Proceedings of the Sixth International Congress of Phonetic Sciences*, Prague, 1967 (Academia, Prague), pp. 563–567.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., and Mandler, R. (**1977**). "On pushing the voice onset-time (VOT) boundary about," Lang Speech **20**, 209–216.

Massaro, D. W., and Oden, G. C. (**1980**). "Evaluation and integration of acoustic features in speech perception," J. Acoust. Soc. Am. **67**, 996–1013.

Menard, S. W. (**1995**). *Applied Logistic Regression Analysis* (Sage Publications, Thousand Oaks, CA).

Miller, J. L. (**1977**). "Nonindependence of feature processing in initial consonants," J. Speech Hear. Res. **20**, 519–528.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Nearey, T. M. (**1997**). "Speech perception as pattern recognition," J. Acoust. Soc. Am. **101**, 3241–3254.

Nittrouer, S., Wilhelmsen, M., Shapley, K., Bodily, K., and Creutz, T. (**2003**). "Two reasons not to bring your children to cocktail parties," J. Acoust. Soc. Am. **113**, 2254.

Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**, 224–230.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Repp, B. (**1979**). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," Lang Speech **22**, 173–189.

Repp, B. (**1983**). "Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization," Speech Commun. **2**, 341–361.

Sawusch, J. R., and Pisoni, D. B. (**1974**). "On the identification of place and voicing features in synthetic stop consonants," J. Phonetics **2**, 181–194.

Soli, S. D., and Arabie, P. (**1979**). "Auditory versus phonetic accounts of observed confusions between consonant phonemes," J. Acoust. Soc. Am. **66**, 46–59.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Stevens, K. N., and Klatt, D. H. (**1974**). "Role of formant transitions in the voiced-voiceless distinction for stops," J. Acoust. Soc. Am. **55**, 653–659.

Strope, B., and Alwan, A. (**1997**). "A model of dynamic auditory perception and its application to robust word recognition," IEEE Trans. Speech Audio Process. **5**, 451–464.

Summerfield, Q., and Haggard, M. (**1977**). "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants," J. Acoust. Soc. Am. **62**, 435–448.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "F0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.

Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index

Petra Larm[a] and Valtteri Hongisto[b]
Laboratory of Ventilation and Acoustics, Finnish Institute of Occupational Health, Lemminkäisenkatu 14–18 B, FIN-20520 Turku, Finland

(Received 5 January 2005; revised 3 October 2005; accepted 9 November 2005)

During the acoustical design of, e.g., auditoria or open-plan offices, it is important to know how speech can be perceived in various parts of the room. Different objective methods have been developed to measure and predict speech intelligibility, and these have been extensively used in various spaces. In this study, two such methods were compared, the speech transmission index (STI) and the speech intelligibility index (SII). Also the simplification of the STI, the room acoustics speech transmission index (RASTI), was considered. These quantities are all based on determining an apparent speech-to-noise ratio on selected frequency bands and summing them using a specific weighting. For comparison, some data were needed on the possible differences of these methods resulting from the calculation scheme and also measuring equipment. Their prediction accuracy was also of interest. Measurements were made in a laboratory having adjustable noise level and absorption, and in a real auditorium. It was found that the measurement equipment, especially the selection of the loudspeaker, can greatly affect the accuracy of the results. The prediction accuracy of the RASTI was found acceptable, if the input values for the prediction are accurately known, even though the studied space was not ideally diffuse. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146112]

PACS number(s): 43.71.Gv, 43.55.Hy, 43.55.Mc [DOS] Pages: 1106–1117

I. INTRODUCTION

When designing different premises, it is often important to know how speech can be perceived in various parts of the room. The demands vary according to the purpose of the premises. In classrooms, the aim is good speech intelligibility throughout the room. On the other hand, in open-plan offices, speech heard from neighboring desks is the most distracting noise source. The more easily the worker can follow the unwanted speech, the more it disturbs concentration[1,2] and, thus, low speech intelligibility between workstations is the aim in open-plan office.

Listening tests are the most natural method to evaluate speech intelligibility. In one form of listening test, the subjects listen to some specific speech material, which has been mixed with disturbances, such as background noise, and register what they have heard, e.g., by writing it down. The speech material may consist of, e.g., short words or syllables, and the results are expressed as the percentage of correctly perceived components. The results depend on the speech material used. Naturally, the perception of words or syllables requires better listening conditions than the perception of sentences. In addition to the characteristics of the speech material, there are many factors that affect the outcome, for instance, the clarity and the accent of the talker, the familiarity of the speech material, and the experience of the test persons. Consequently, a listening test does not provide an absolute value of intelligibility of a certain listening condition but rather a method to compare different conditions with that particular test setup. As the arrangement of listening tests is a rather laborious process, several objective methods have been developed to measure and predict speech intelligibility.

In this study, two physical methods to measure speech intelligibility are evaluated, namely the *speech intelligibility index (SII)* and the *speech transmission index (STI)*. The *room acoustics speech transmission index* (also called *RApid Speech Transmission Index) (RASTI)*, which is a simplified method of STI, is also considered. These are standardized methods, SII by ANSI S3.5-1997,[3] and STI and RASTI by IEC 60268-16, 3rd ed. (2003).[4]

The SII is a revised version of the earlier articulation index (AI),[5,6] which was a simplification of the methods presented in Ref. 7. The AI was intended for measuring speech intelligibility through communication channels with distortions in the frequency domain including interfering noise and band-pass limiting. Time domain distortions, like reverberation, were not accounted for. In ANSI S3.5-1997, measurement procedures have been improved and the present SII is capable of measuring the speech transmission also in reverberant fields. The communication system is assumed to be linear. Thus, the SII does not take into account nonlinear effects.

The STI has been developed for the evaluation of speech intelligibility both in direct communication situations and in situations where the speech is amplified by, e.g., a public address system. The STI takes into account various distor-

[a]Electronic mail: petra.larm@ttl.fi
[b]Electronic mail: valtteri.hongisto@ttl.fi

tions, e.g., reverberation, echoes, peak clipping, and other nonlinear distortion and interfering noise. Frequency shifts or multiplication, or systems with vocoders, are not accounted for. The RASTI is a simplified screening method of the STI and it is meant for room acoustics and direct communication situations. Thus, the RASTI is limited to situations where no band-pass limiting or nonlinear distortion exists. Reverberation and background noise should not be strongly frequency dependent, and the background noise should not be impulsive or include clear tones. There are also other simplifications of the STI: STITEL (for measuring telecommunication channels) and STIPA (for measuring public address systems). These are presented in IEC 60268-16 (2003) but they were not considered in this study.

In IEC 60268-16 (2003) standard, the method is called the STI_r since it is a revised version of the original STI, which is described in Ref. 8. The original STI has been further improved and it takes into account such effects as noncontiguous frequency transfer and severe band-pass limitation. The auditory masking and absolute hearing threshold are also considered. Moreover, male and female speakers are treated separately. In this study, a simplified STI version described in Ref. 9 was used, as it was implemented in the measuring software version used at that time. This is the original version except that it does not take into account the masking effects. However, it is still a valid method for assessing speech intelligibility in rooms where noncontiguous frequency transfer or band-pass limiting do not cause problems, or the speaking levels are not extremely low.[9,10]

A detailed description of the development and the use of the STI and RASTI can be found from Ref. 11. Development of the STI and RASTI was based on finding the best fit with subjective speech intelligibility using specific speech material. It has been validated that the STI and RASTI correlate well with subjective speech intelligibility in different listening conditions and in many languages.[10,12]

The STI and RASTI have been extensively used for measuring and predicting speech intelligibility in various spaces, e.g., in classrooms,[13,14] churches,[15,16] courtrooms,[17] cloisters,[18] railway stations,[19,20] and open-plan offices.[1,21] Different measuring devices have been developed for measuring the STI and RASTI. For comparison, some data were needed on their performance and possible discrepancies in various speech-to-noise ratios and reverberation conditions. For example, the choice of the measurement loudspeaker is of interest, because often the loudspeaker available is used, despite being less than optimal for the measurement. The differences between the STI and RASTI are also of interest, as it has been observed that the RASTI does not necessarily depict the speech intelligibility in a reliable way, if the reverberation or background noise conditions are not smooth in the frequency bands considered.[22]

In the IEC 60268-16 standard, there are data of the speech intelligibility scores on different speech material for the STI. In the ANSI S3.5-1997 standard, there are given band importance functions, which are used in the SII calculation and which characterize the relative significance of each frequency band to speech intelligibility depending on the type of the speech material. Comparison between the STI and SII with these data is, however, difficult. What is the actual difference between these single-number quantities when both are measured in the same conditions?

In the prediction of the STI, a diffuse sound field is assumed.[23] The accuracy of the prediction in nondiffuse rooms has not yet been studied extensively. What is the difference between the predicted and the measured (RA)STI values in an arbitrary real nondiffuse space, when the input values for the prediction are accurately known?

The scope of this study was to compare the STI, RASTI, and SII methods and their measuring equipment performance in a laboratory and in a real auditorium. It was assumed that the STI and RASTI are valid quantities to depict speech intelligibility in their scope of application. Thus, no subjective listener tests were carried out. In the laboratory, reverberation and speech-to-noise ratios were adjusted, but the distance between the assumed speaker and the listener was fixed. In the auditorium, reverberation and background noise were quite even throughout the room, but the distances between the speaker and the listener, and the orientation of the speaker to the listener, varied. The communication was assumed to be direct, i.e., person-to-person without any speech amplification. Echoes, reverberation, and background noise were the only disturbances mixing with the speech.

For this study, there were four main tasks. The results and discussion are outlined in these four categories.

Comparison between the STI and RASTI. Measured STI and RASTI values were compared to see how well the RASTI correlates with the STI, and whether the RASTI is a feasible simplification of the STI.

Comparison between different RASTI measuring procedures. RASTI values were measured with two devices to detect any possible variation between them caused by both the different measuring procedure and the measuring equipment. The first procedure used a random test signal and the second was based on a *maximum length sequence* (*MLS*) analysis.

Comparison between the STI and SII. To see if there are essential differences between the STI and SII, the values were measured and compared.

Comparison between the measured and the predicted (RA)STI. RASTI and STI values were predicted using the measured reverberation times and speech and noise levels, and then compared with the measured ones.

II. METHODS

A. Modulation transfer function (MTF)

The actual information of speech lies in the *modulations of the speech intensity envelope* generated by the words and syllables of running speech. For perfect intelligibility, the original speech should preserve all its intensity envelope modulations on the way from the speaker to the listener. In direct communication between people, the room reverberation and background noise affect the speech modulations in a disruptive way. In both cases, the relative depth of the speech intensity envelope modulations is reduced. With the *modulation transfer function (MTF)* method,[23] the modulation reduction can be determined.

The MTF can be measured using a test signal composed of band-limited random noise, whose spectrum and level correspond to those of human speech, and whose intensity is sine-modulated with the modulation frequencies relevant to speech. This test signal is played through a loudspeaker positioned at the location of the speaker, having directivity similar to that of the human mouth, and measured with a microphone at the location of the listener. The transmitted and received signals are compared and the relative modulation reduction of the intensity modulations is detected. To get the MTF, a *modulation reduction factor m*, which is defined [Eq. (1)] as the ratio of the modulation indices of the received signal m_r and the original test signal m_t, is determined with the modulation frequencies F used to modulate the test signal. The MTFs are measured in the frequency range significant to speech, i.e., octave bands 125–8000 Hz, with modulation frequencies roughly from 0.5 to 16 Hz:

$$\mathrm{MTF} = m(F) = \frac{m_r}{m_t}. \tag{1}$$

The modulation reduction factor m at modulation frequency F caused by reverberation can also be measured using the impulse response of that special speaker-listener-pair with Eq. (2):

$$m(F) = \frac{\left| \int_0^\infty e^{-i2\pi Ft} h^2(t)\, dt \right|}{\int_0^\infty h^2(t)\, dt}, \tag{2}$$

where t is time and $h(t)$ is the (noiseless) impulse response of that exact speaker-listener-pair. Background noise causes an additional modulation reduction factor m_n, which is of the form

$$m_n = \frac{1}{1 + 10^{-\mathrm{SNR}/10}}. \tag{3}$$

SNR is called the speech-to-noise ratio, which is defined as the difference between the speech level and the noise level. Speech level is considered as the total level at the listener's location, i.e., the sum of the direct sound, first reflections, and reverberant speech. In general, when both reverberation and background noise are present, the modulation reduction factor m is obtained as a product of Eq. (2) and Eq. (3).

Measuring the MTF using Eqs. (2) and (3) together requires two separate measurements, i.e., the noiseless impulse response of the speaker-listener pair and the speech-to-noise ratio. It is assumed that the speech and the noise levels can be measured independently, which means that there should not be any speech-dependent noise. In rooms where electrical systems are not used, this is the case, since in the MTF method, the reverberant part of the speech is not considered as noise but as a part of the speech level. With the *maximum length sequence* (*MLS*) technique, it is possible to get an accurate approximation of the noiseless impulse response using averaging. With the MLS, it is also possible to measure simultaneously the modulation reduction caused by reverberation and background noise, if the MLS stimulus signal is weighted to speechlike and no averaging is done.

The MTF can be predicted if some characteristics of the room are known. Assuming that the only affecting disturbance is reverberation, and that the reverberation is of an exponential nature, Eq. (2) can be simplified into

$$m_{\mathrm{rev}}(F) = \frac{1}{\sqrt{1 + (2\pi FT/13.8)^2}}, \tag{4}$$

where T is the early decay reverberation time (EDT) at the listener point. It is also assumed that the listener is in the far field of the speaker's speech, so that the direct sound can be disregarded.

If the speaker is close to the listener, the enhancement of the direct sound must be taken into account. A more complicated form of the modulation reduction factor $m_{\mathrm{rev},d}$ can then be derived from Eq. (2) to yield Eq. (5):[23]

$$m_{\mathrm{rev},d}(F) = \frac{\sqrt{(q_s q_l)^2 (r_c/r)^4 + \{1/[1 + (2\pi FT/13.8)^2]\}(2 q_s q_l (r_c/r)^2 + 1)}}{q_s q_l (r_c/r)^2 + 1}. \tag{5}$$

Now the MTF depends not only on the reverberation time T and modulation frequency F but also on the room's critical radius r_c, the speaker-listener distance r, and the directivity factors of the speaker q_s and the hearing of the listener q_l. The room's critical radius is the distance from the source, where the strengths of the direct and the reverberant sound field are equal. This can be calculated from[24] $r_c = 0.1\sqrt{V/\pi T}$, where V is the volume of the room.

In order to evaluate speech intelligibility, the measured or the predicted MTFs can be converted into one figure that characterizes speech intelligibility. The STI, RASTI, and SII are such measures. The relation between the actual speech intelligibility and the physical measure is a monotonically increasing function whose exact form depends on the speech material used. Relations between speech intelligibility and these physical measures have been derived for different speech materials, e.g., phonetically balanced words, equally balanced consonant-vocal-consonant words, or sentences.[3,4]

The directions for the calculations of all these three mea-

sures can be found from the above-mentioned standards (and Ref. 9 for the STI version used in this study), but they are also presented below for easier comparison.

B. STI calculation

The MTFs are measured in seven octave frequency bands with middle frequencies, f_i, from 125 to 8000 Hz. Every MTF is measured with 14 modulation frequencies, F_j, from 0.63 to 12.5 Hz in one-third octave bands. As a result, 98 modulation reduction factors are obtained.

Every modulation reduction factor is transferred into a so-called apparent signal-to-noise ratio $\mathrm{SNR}_{\mathrm{app}}$ [Eq. (6)]. This expresses the signal-to-noise ratio that would alone cause the measured modulation reduction, if it were the only factor distorting the speech signal:

$$\mathrm{SNR}_{\mathrm{app}_{i,j}} = 10 \log \left(\frac{m(f_i,F_j)}{1-m(f_i,F_j)} \right) \mathrm{dB}. \tag{6}$$

The dynamic range of speech is assumed to 30 dB: the $\mathrm{SNR}_{\mathrm{app}_{i,j}}$ values are limited to the range of −15–15 dB. In every frequency band, f_i, the 14 $\mathrm{SNR}_{app_{i,j}}$ values are averaged. The frequency band specific STI_i is acquired by converting the averaged $\overline{\mathrm{SNR}}_{\mathrm{app}_i}$ into the range of 0–1. From the seven frequency-band-specific STI_i values a weighted summation is calculated. The weighting coefficients, w_i, for the seven frequency bands are 0.13, 0.14, 0.11, 0.12, 0.19, 0.17, and 0.14, respectively:[9]

$$\mathrm{STI} = \sum_{i=1}^{7} w_i \, \mathrm{STI}_i = \sum_{i=1}^{7} w_i \frac{\overline{\mathrm{SNR}}_{\mathrm{app}_i} + 15}{30}$$

$$= \sum_{i=1}^{7} w_i \cdot \frac{\left(\frac{1}{14} \sum_{j=1}^{14} \mathrm{SNR}_{\mathrm{app}_{i,j}} \right) + 15}{30}. \tag{7}$$

According to (7), the STI will be zero, if the $\overline{\mathrm{SNR}}_{\mathrm{app}_i}$ is −15 dB. If there is nothing to distort the speech signal, the $\overline{\mathrm{SNR}}_{\mathrm{app}_i}$ will be 15 dB and the STI will approach to unity.

C. RASTI calculation

The RASTI is a simplified version of the STI. It is calculated using only the 500- and 2000-Hz octave bands. In the 500-Hz band the modulation reduction factor is calculated with four modulation frequencies F_j (1, 2, 4, and 8 Hz) and in the 2000-Hz band with five modulation frequencies (0.7, 1.4, 2.8, 5.6, and 11.2 Hz). After transferring them into apparent signal-to-noise ratios, all nine $\mathrm{SNR}_{\mathrm{app}}$ values are averaged. Then the weighting factor for the 500-Hz frequency band can be thought of as 4/9 and for the 2000 Hz band as 5/9. Finally, the RASTI value is acquired by converting the averaged $\overline{\mathrm{SNR}}_{\mathrm{app}}$ to the range of 0–1:

$$\mathrm{RASTI} = \frac{4}{9}\mathrm{STI}_{500\mathrm{Hz}} + \frac{5}{9}\mathrm{STI}_{2000\mathrm{Hz}}$$

$$= \frac{\left(\frac{1}{9} \sum_{j=1}^{9} \mathrm{SNR}_{\mathrm{app}_j} \right) + 15}{30}. \tag{8}$$

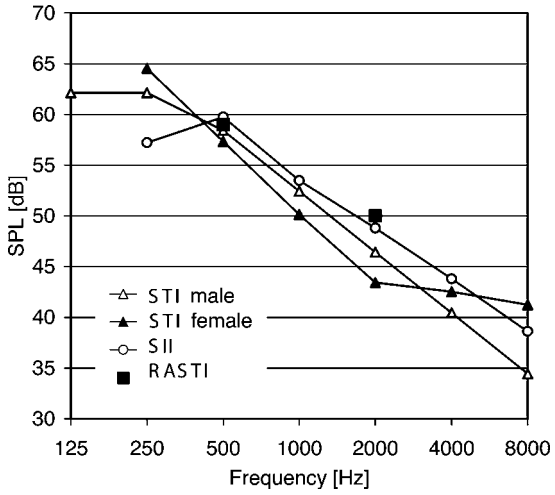


FIG. 1. The speech spectra. The SII spectrum was used in this study for both SII and STI calculation (free field, 1 m in front of the speaker).

D. SII calculation

The SII calculation can be done using several frequency band divisions. In this study, the octave frequency bands were used. The calculation was done in bands with middle frequencies, f_i, from 250 to 8000 Hz. The SII can be calculated from

$$\mathrm{SII} = \sum_{i=1}^{6} L_i \cdot I_i \cdot \frac{(E_i' - D_i) + 15}{30}. \tag{9}$$

E_i' is called the spectrum level of equivalent speech and D_i is called the equivalent disturbance spectrum level in the octave band with middle frequency, f_i. The difference $(E_i' - D_i)$ is limited to the range of −15–15 dB.

I_i's are the band-weighting coefficients or band importance functions, as they are called in the standard. They are for the six frequency bands 0.0617, 0.1671, 0.2373, 0.2648, 0.2142, and 0.0549, respectively. These values are the default, and they apply to average speech. When predicting speech intelligibility for a specific speech material, band importance functions are listed in the standard for a number of speech tests. L_i is a level distortion factor, which accounts for the decrease in speech intelligibility, if the speech level is too high. It is calculated from

$$L_i = 1 - \frac{E_i' - U_i - 10}{160}, \tag{10}$$

where U_i is the standard speech spectrum level (Fig. 1) for the frequency band f_i. The values of L_i are limited to the range of 0–1.

The equivalent disturbance level D_i is obtained by choosing the larger value from the equivalent internal noise spectrum level X_i' (given in the standard[3]) and the equivalent noise spectrum level N_i'. The use of the equivalent internal noise spectrum level takes into account the fact that, in silent conditions, the fictitious internal noise in one's ear masks the sounds coming from outside the ear.

In the SII calculation, the equivalent speech and noise levels are assumed to be measured in free field conditions. This means that the total speech level contains no reverber-

ant speech. It is also assumed that the speech and noise sources can be measured in the absence of each other. According to the standard, in low reverberant conditions, where reverberation does not reduce speech intelligibility and the noise does not depend on the speech, it is possible to measure the equivalent speech spectrum level E_i' and the equivalent noise spectrum level N_i' directly. In reverberant fields, however, it is not possible to measure the speech level without the reverberant part that is considered as noise. Then the equivalent speech and noise spectrum levels can be determined using the MTF method. The apparent signal-to-noise ratio $\mathrm{SNR}_{\mathrm{app}_{i,j}}$ [Eq. (6)] should be determined with nine modulation frequencies F_j: 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 6.0, 8.0, and 16.0 Hz. In every frequency band, f_i, the nine $\mathrm{SNR}_{\mathrm{app}_{i,j}}$ values are limited to the range of −15–15 dB and averaged. The equivalent noise and speech spectrum levels are obtained from Eqs. (11) and (12), where P_i is the combined speech and noise level. It is measured at the location of the listener in the presence of the ambient noise using a speech test signal made from random noise, shaped to correspond to the standard speech spectrum and played at the location of the speaker:

$$N_i' = 10 \log \left(\frac{10^{P_i/10}}{1 + 10^{\overline{\mathrm{SNR}}_{\mathrm{app}_i}/10}} \right), \tag{11}$$

$$E_i' = \overline{\mathrm{SNR}}_{\mathrm{app}_i} + N_i'. \tag{12}$$

E. Differences between the STI, RASTI, and SII

The SII and STI calculation procedures are quite similar. Both these parameters are based on determining an apparent speech-to-noise ratio on selected frequency bands and summing them using a specific weighting. Differences between these indices can arise from the fact that the SII takes into account the extreme cases in speech-to-noise ratio, which the STI version used here does not. The degrading effect of a high presentation level of speech on speech intelligibility has, in practice, no relevance at normal speech levels. The masking of faint sounds by the internal noise in the ear is of relevance predominantly on the highest frequency band 8000 Hz when the speech-to-noise ratio is very high. The apparent signal-to-noise ratio determination is basically the same in both methods, when reverberant conditions are assumed. Then both methods utilize the MTF method, with only the modulation frequencies being slightly different. More important than these factors are the differences in the band weighting coefficients. Their differences are due to the different speech material used in the development of the methods. In addition, the STI calculation is done with seven octave bands from 125 to 8000 Hz, whereas in the SII calculation the 125-Hz band is excluded.

STI and RASTI calculation procedures are the same except that the RASTI uses only two frequency bands and nine modulation frequencies instead of seven frequency bands and 14 modulation frequencies. As RASTI is calculated using only two frequency bands, it is necessary that the background noise and reverberation are not strongly frequency dependent in the room being considered. Moreover, complicated MTF may not be detected reliably, because of the reduced number of modulation frequencies.

The assumed speech spectrum levels are different for the methods considered, and these are presented in Fig. 1. The levels are given in a free field 1 m in front of the speaker. The IEC 60268-16 (2003) standard gives separate speech spectrum levels for male and female speakers. In this study, the ANSI S3.5-1997 spectrum was used for the evaluation of both the STI and SII for simplicity. The total A-weighted speech level was then 59 dB. For the STI calculation, the 125-Hz band level was assumed to be the same as the 250-Hz band level.

TABLE I. Early decay times of the reverberation room (1–6) and the semi-anechoic room (7).

	f (Hz)						
Measurement	125	250	500	1000	2000	4000	8000
1	1.97	2.07	3.09	2.51	1.95	1.14	0.65
2	2.04	1.95	2.24	1.87	1.73	1.04	0.64
3	1.60	1.20	1.24	1.19	1.28	0.97	0.61
4	1.60	1.11	0.87	0.95	0.88	0.79	0.50
5	0.87	0.62	0.44	0.51	0.51	0.41	0.29
6	0.65	0.49	0.26	0.26	0.31	0.32	0.23
7	0.31	0.12	0.11	0.05	0.06	0.06	0.07

III. EXPERIMENTAL METHODS

A. Laboratory measurements

In the laboratory, STI, RASTI, and SII values were measured in different reverberation and background noise configurations. The speech-to-noise ratios varied between −15 and 15 dB with 5-dB steps. The early decay reverberation time varied at 500 Hz from 0.1 to 3.1 s (Table I). Overall, seven measurement series were carried out, six of them in a reverberation room (volume 113 m^3) and one in a semi-anechoic room (volume 77 m^3). The reverberation conditions of the reverberation room were varied by the addition of absorption material and background noise in both rooms by playing band-limited pink noise through a loudspeaker.

The walls, the ceiling, and the floor of the reverberation room were highly reflective. Three diffusers were mounted on the ceiling to make the sound field more diffuse. The room reverberation was varied by installing 45–100-mm-thick mineral wool plates of different sizes in the room. The positioning of the absorption material was made as even as possible to get smoother reverberation, although the upper parts of the walls and the ceiling were left reflective for practical reasons. The semi-anechoic room had a 1100-mm suspended ceiling of 40-mm-thick mineral wool. The walls of the room were lined with 300-mm-thick mineral wool. The floor was painted concrete.

EDTs of the seven measurement series are presented in Table I. Measurements 1–6 were made in the reverberation room and measurement 7 in the semi-anechoic room. In measurement 1, the reverberation room was empty of any absorption material.

The measurement loudspeaker was mounted on a tripod at a height of 1.2 m near the corner of the room. The microphone was mounted at the same height and 2.4 m away from the loudspeaker, pointing towards it. The loudspeaker was turned 90° so that it pointed to the wall in order to attenuate the direct sound. The background noise was adjusted to speechlike so that the speech-to-noise ratio was essentially the same in every octave band at the microphone location. The noise was generated and weighted with a Behringer Ultra-Curve Pro DSP 8024 frequency filter and played through a Fostex SPA 100 loudspeaker.

The speech levels were determined and background noise levels adjusted every time that absorption material was added. The "speech" was composed of octave band filtered pink noise. The pink noise was filtered so that its levels corresponded to the ANSI standard speech spectrum measured 1 m in front of the speaker in a free sound field. The speech test signal was played with a Genelec 1029A loudspeaker and the levels were measured at the microphone location with a Brüel & Kjær $\frac{1}{2}$-in. 4190 microphone and a Norsonic Real-Time Analyser RTA 840. Thereafter, the background noise levels were adjusted to correspond to the measured speech levels at the microphone location.

The STI, RASTI, and EDT values were measured with a PC and WinMLS2000 Pro/Room Acoustics program in each reverberation configuration. WinMLS2000 determines the STI values using the STI version described in Ref. 9. MTFs are acquired using the MLS averaging technique, so that, in practice, the obtained impulse response estimates an ideal situation without background noise as in Eq. (2). To calculate the STI with different speech-to-noise ratios, the total speech and noise levels were measured. The MLS signal sent by the program was played through a Genelec 1029A loudspeaker and the response was recorded with a B&K 4190 microphone. The microphone was connected to a Norsonic RTA 840 for microphone amplification and for measurement of the speech and noise levels. The reverberant speech levels could be measured directly using the speech test signal because the speech levels were high enough compared with the background noise in the laboratory room (A-weighted level 29 dB).

RASTI values were also measured with a B&K Speech Transmission Meter type 3361, which consists of a test signal transmitter and a receiver. As the test signal, the device uses random pink noise, which is modulated by the nine RASTI modulation frequencies simultaneously. The averaging time was 32 s. The loudspeaker is supposed to have directivity properties close to those of the human head.[25] Compared with the Genelec 1029A, the RASTI transmitter is slightly more omnidirectional at the 500-Hz frequency band. At the 2000-Hz frequency band, it directs sound more upward, i.e., to an approximately 50° elevation, whereas with the Genelec 0° elevation is the dominant direction. The RASTI measured with WinMLS is later referred to as the MLS method and the RASTI measured with the B&K 3361 as the modulated test signal method. The parameters acquired with these two procedures are called the $\mathrm{RASTI}_{\mathrm{MLS}}$ and $\mathrm{RASTI}_{\mathrm{mod}}$, respectively.

The SII values were calculated using the MTFs obtained with WinMLS and, thus, with the same modulation frequencies as used in the STI calculation, but the error is quite small. This will be discussed later in more detail. For the SII calculation, the combined speech and noise spectrum levels were also measured, i.e., the sound pressure levels at the receiver location, when the speech and noise sources were both running simultaneously.

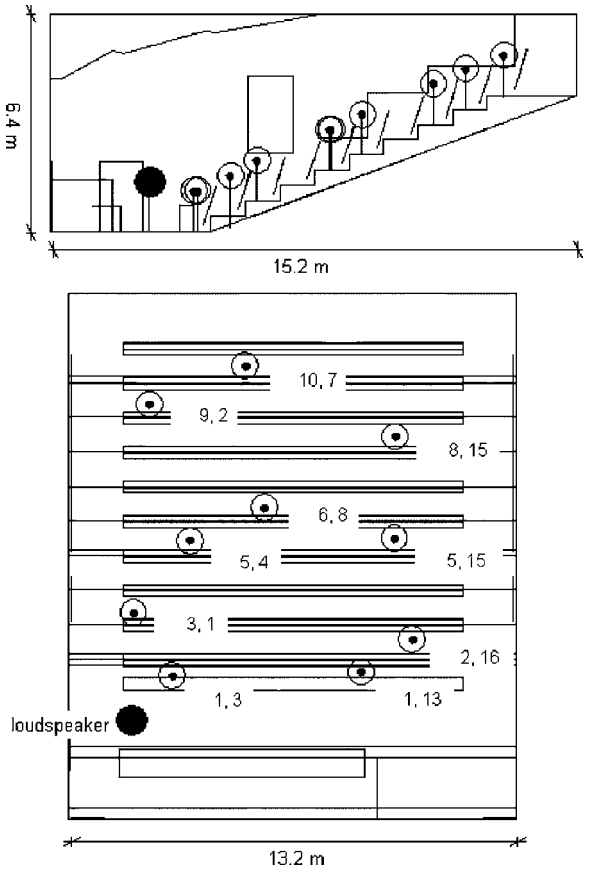


FIG. 2. The auditorium. The first number in receiver locations gives the row number and the second the seat line number.

B. Auditorium measurements

The auditorium had seats for 180 listeners, ten rows and 18 seats in every row. The volume of the auditorium was approximately 1050 m^3. The walls were of concrete and the floor was covered with a soft carpet. Some absorbent material was hung on the walls and the suspended ceiling was also covered with the same material.

The STI, RASTI, and SII were measured in ten different locations located in the audience area (Fig. 2). The measurements were made in two background noise conditions, first with the air conditioning off and then with it on. The A-weighted background noise levels at the receiver locations were then 27–30 dB and 39–46 dB, respectively. The speech-to-noise ratios varied according to the measurement point. At lower background noise conditions, the speech-to-noise ratios were more than 14 dB at every frequency band. At higher background noise conditions, the average speech-to-noise ratios were 4.1, 5.6, 13.9, 11.1, 8.8, 7.8, and 9.2 for the frequency bands 125–8000 Hz, respectively.

The measurements were made with one sound source location and ten receiver locations (Fig. 2). The sound source was mounted on a tripod at a height of 1.55 m in the front part of the room, 1.5 m from the side wall pointing towards the opposite corner at the back of the seating area. Two sound sources were used in turn, a Genelec 1029A loudspeaker and a B&K OmniPower 4296 omnidirectional loudspeaker (below: directional loudspeaker and omnidirectional loudspeaker, respectively). Again, the STI and RASTI were measured with WinMLS and the RASTI also with the B&K 3361. SII values were calculated. The background noise, the speech, and the combined speech and noise spectra were measured with a B&K Investigator 2260A with a B&K 4189 microphone. In the impulse response determination, a B&K 2231 sound level meter with a B&K 4165 microphone connected to a PC was used with WinMLS.

The auditorium is shown in Fig. 2. The receiver locations were given two numbers, the first one the row number (calculated from the front) and the second the seat line number (calculated from the left).

In order to diminish the effect of background noise, the speech levels were measured by playing pink noise with a B&K 4296 loudspeaker, with known sound power levels. With the known sound power levels, the sound pressure levels at 1 m distant from the loudspeaker in the free field can be estimated, assuming an omnidirectional radiation.[24] These levels were compared with the measured sound pressure levels at the receiver locations, and attenuations were calculated. Then, the observed attenuations were added to the ANSI standard speech spectrum levels (given at a distance of 1 m in free field). The combined speech and noise spectrum levels were determined applying the same speech test signal procedure as in the laboratory measurements but with the omnidirectional loudspeaker.

There was a strong flutter echo between the side walls of the auditorium. Flutter echo affects the MTF, if the flutter echo repetition frequency is an integral multiple of some modulation frequency.[4] The width of the room was 13.2 m so the flutter echo frequency was about 13 Hz. As the largest modulation frequency is 12.5 Hz, the flutter may have a small effect on the largest modulation frequency.

EDTs measured with WinMLS and the omnidirectional loudspeaker are presented in Table II. The EDTs obtained with the omnidirectional sound source were in most cases somewhat higher than those measured with the directional loudspeaker.

TABLE II. Early decay times of the auditorium measured at the ten receiver locations (see Fig. 2) using the omnidirectional sound source.

	f (Hz)						
Location	125	250	500	1000	2000	4000	8000
1,3	1.33	0.70	0.60	0.31	0.24	0.24	0.16
1,13	1.14	1.04	0.66	0.51	0.54	0.46	0.32
2,16	1.60	1.14	0.63	0.65	0.42	0.42	0.36
3,1	1.28	1.13	0.76	0.53	0.46	0.39	0.23
5,4	1.64	1.03	0.76	0.61	0.57	0.55	0.39
5,15	1.52	0.99	0.72	0.66	0.59	0.63	0.41
6,8	1.37	1.07	0.75	0.51	0.56	0.51	0.40
8,15	1.08	1.23	0.79	0.63	0.56	0.43	0.29
9,2	1.56	1.34	0.84	0.53	0.67	0.49	0.43
10,7	1.23	1.11	0.76	0.66	0.75	0.52	0.41

IV. RESULTS

A. Comparison between the STI and RASTI

The differences between STI and RASTI values as a function of the speech-to-noise ratio, measured with the MLS method in the laboratory rooms, are presented in Fig. 3. Measurements 1–6 were made in the reverberation room and measurement 7 in the semi-anechoic room. The speech-to-noise ratio was essentially the same for every frequency band.

The differences between STI and RASTI values measured with the MLS method in the auditorium are presented in Fig. 4. In the lower background noise conditions, the RASTI values were on average 0.01 higher than the STI values, when the omnidirectional loudspeaker was used, and on average 0.02 higher than the STI values, when the directional loudspeaker was used. In the higher background noise conditions, the RASTI values were on average 0.04 higher than the STI values (omnidirectional loudspeaker).

B. Comparison between different RASTI methods

The differences between the $RASTI_{mod}$ and $RASTI_{MLS}$ values as a function of the speech-to-noise ratio, measured in the laboratory rooms, are presented in Fig. 5. Respectively, the differences between these RASTI methods measured in the different auditorium locations are presented in Fig. 6.

Both in the laboratory and in the auditorium, the background noise was of a steady nature. The standard deviations of $RASTI_{mod}$ values were below 0.03. The deviations were slightly higher at $RASTI_{mod}$ values below 0.25 than above it. The $RASTI_{mod}$ value was obtained as an average of four measurements in every reverberation and background noise

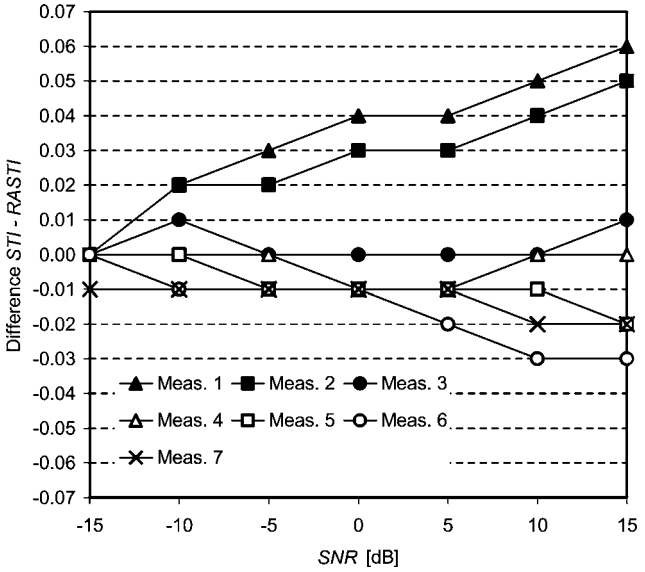


FIG. 3. The difference between STI and RASTI values measured with the MLS method in the laboratory rooms. Measurement 1 corresponds to the longest reverberation time and measurement 7 the shortest (see Table I).

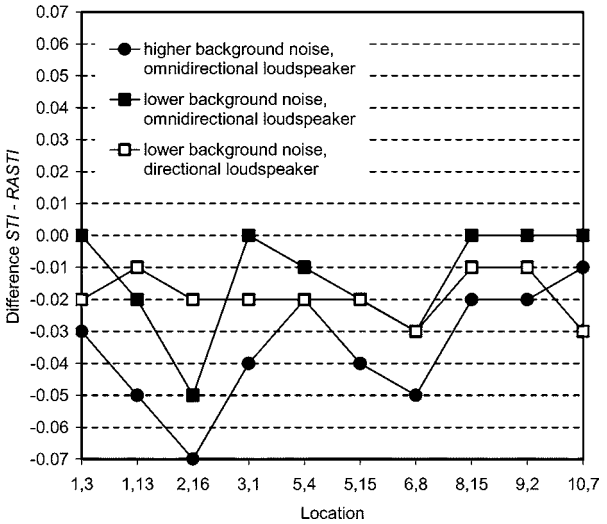


FIG. 4. The difference between STI and RASTI values measured with the MLS method in the auditorium. The measurements were made with the air conditioning on [L_n=39–46 dB(A)] and off [L_n=27–30 dB(A)]. With the lower background noise, two different loudspeakers were tested. For location labels, see Fig. 2.

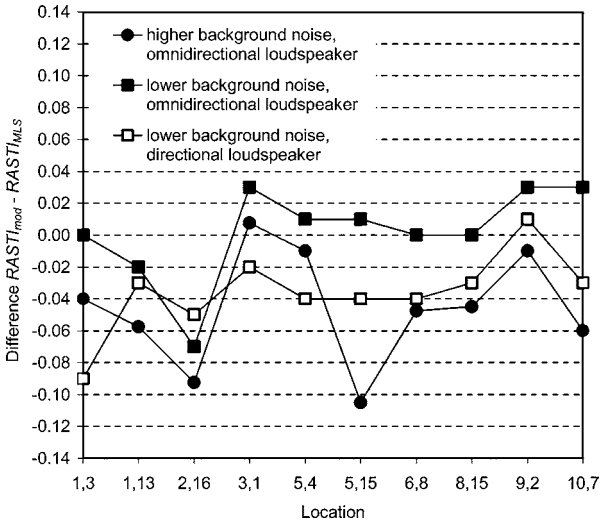


FIG. 6. The difference between the RASTI$_{mod}$ and RASTI$_{MLS}$ in the auditorium. The measurements were made with the air conditioning on [L_n=39–46 dB(A)] and off [L_n=27–30 dB(A)]. With the lower background noise, two different loudspeakers were tested in the MLS measurements. For location labels, see Fig. 2.

configuration. With the MLS method, the deviation between two successive STI or RASTI measurements was usually 0.01 or less.

C. Comparison between the STI and SII

The differences between the STI and SII values as a function of the speech-to-noise ratio, measured in the laboratory rooms, are presented in Fig. 7. The differences between these methods measured in the different auditorium locations are presented in Fig. 8.

D. Comparison between measured and predicted (RA)STI

Some MTFs measured in the 125-, 500-, and 8000-Hz frequency bands in the auditorium are presented in Fig. 9. The measurements were made in location 6, 8. Also shown are the predicted (direct sound taken into account) MTFs, which are calculated using Eqs. (3) and (5) and the measured speech-to-noise ratios and EDTs (obtained with the omnidirectional loudspeaker). The directivity factor of the speaker was assumed to be q_s=2 and the directivity factor of the listener q_l=1.5.[23]

With the calculated MTFs, the STI and RASTI can be predicted using the calculation procedures described in methods. In Fig. 10, some of the measured and predicted RASTI$_{mod}$ values in the 500-Hz frequency band are presented as a function of the speech-to-noise ratio. The measurements were made in the laboratory rooms. In the prediction, the direct sound was again taken into account.

The predicted and the measured STI and RASTI values for the different locations in the auditorium are presented in Table III. The A-weighted background noise was 27–30 dB. The prediction was made both taking the direct sound into

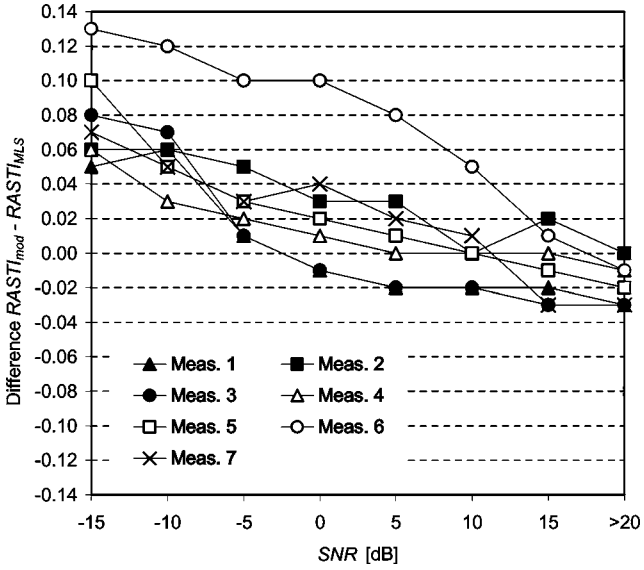


FIG. 5. The difference between RASTI values measured with the modulated test signal (RASTI$_{mod}$) and the MLS method (RASTI$_{MLS}$) in the laboratory room.

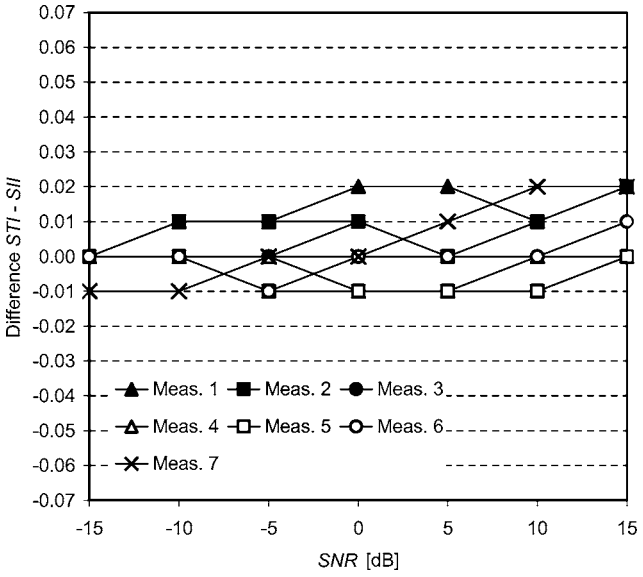


FIG. 7. The difference between the STI and SII in the laboratory rooms.

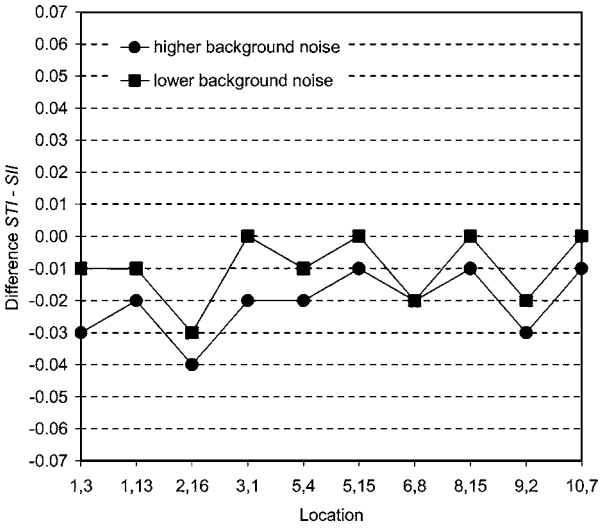


FIG. 8. The difference between the STI and SII in the auditorium.

account [Eq. (5)] and disregarding it [Eq. (4)]. The calculation was also made separately for the two loudspeakers, because the EDTs obtained with them were different. The speech level determination was made only with the omnidirectional loudspeaker. Generally, the predicted values were slightly closer to the measured ones when the calculation was made using the reverberation times acquired with the omnidirectional loudspeaker.

In Fig. 11, the measured RASTI values are compared with the predicted ones.

V. DISCUSSION

A. Comparison between the STI and RASTI

In the laboratory (Fig. 3), at low speech-to-noise ratios, the differences between the STI and RASTI were, as expected, small, because the speech-to-noise ratio was the main parameter affecting the STI, and it was the same in every frequency band. As the speech-to-noise ratio increased and the effect of reverberation became more significant, the differences also increased. The differences between STI and

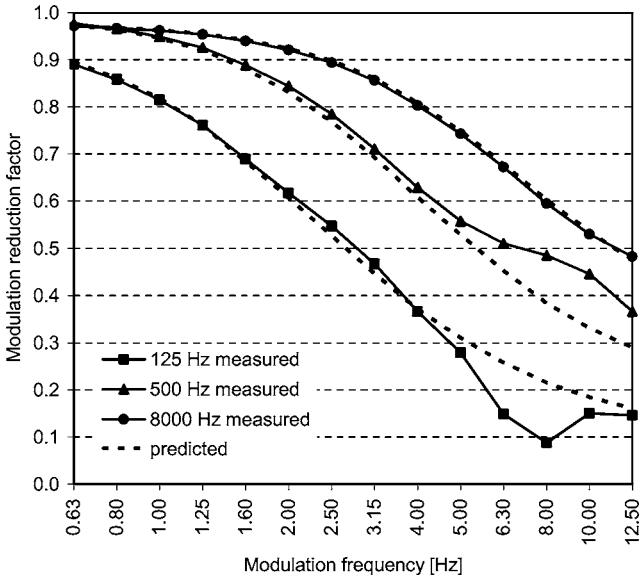


FIG. 9. Some of the measured and predicted MTFs (location 6, 8 in auditorium).

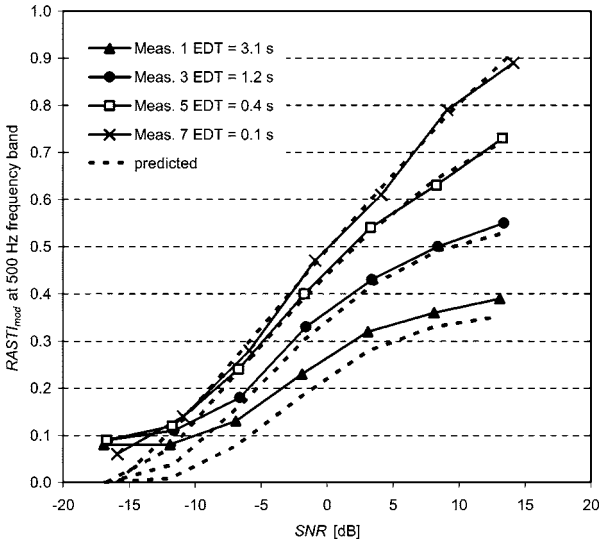


FIG. 10. Measured and predicted $RASTI_{mod}$ values in the 500-Hz frequency band as a function of the speech-to-noise ratio. The measurements were made in the laboratory rooms.

RASTI values were largest in measurements 1 and 2, where reverberation was quite different in 500- and 2000-Hz bands compared with other bands (see Table I).

Even though the differences between the STI and RASTI were not substantial in this particular study, there are certainly situations when it is better to use the STI. The reverberation time and background noise are rarely smooth enough, and, anyway, the conditions should always be checked.

B. Comparison between different RASTI methods

The randomness of the modulated test signal results in the fact that the RASTI values measured with the modulated test signal rarely attain the minimum (0) and maximum (1) values. Thus, the values obtained with the modulated test signal method were usually higher than those obtained with the MLS method with very low RASTI values and vice versa with high RASTI values.

The transmitter levels of the modulated test signal are slightly different in the frequency bands considered from the ANSI standard speech spectrum levels used in the MLS method. The RASTI transmitter level is 0.8 dB lower than the ANSI level in the 500-Hz frequency band and 1.2 dB higher in the 2000-Hz frequency band. The effect of the differences in the assumed speech spectrum can be estimated using the measured reverberation times, the speech-to-noise ratios, and Eqs. (4), (6), and (8). In this laboratory case, the RASTI values acquired with the modulated test signal would be at most 0.02 higher than those acquired with the MLS method, if the speech spectrum discrepancy were the only source of error. The errors caused by the randomness of the test signal and the discrepancy of the speech spectrum are, thus, summed together at low speech-to-noise ratios and compensate for each other at high speech-to-noise ratios in this laboratory case. The observed behavior in Fig. 5 conforms to this.

These two effects, however, only partly explain the differences between the RASTI methods measured in the labo-

TABLE III. Predicted and measured STI and RASTI in the auditorium. "Direct" means that direct sound has been taken into account in the prediction.

			Location									
	Loudspeaker	Parameter	1,3	1,13	2,16	3,1	5,4	5,15	6,8	8,15	9,2	10,7
Predicted	omnidirectional	STI	0.77	0.69	0.68	0.70	0.66	0.65	0.67	0.67	0.65	0.65
		STI direct	0.93	0.71	0.69	0.79	0.69	0.66	0.68	0.68	0.66	0.65
		RASTI[a]	0.79	0.70	0.73	0.70	0.68	0.68	0.68	0.68	0.65	0.65
		RASTI direct[a]	0.95	0.72	0.74	0.80	0.71	0.69	0.70	0.68	0.66	0.65
	directional	STI	0.88	0.67	0.66	0.71	0.65	0.68	0.66	0.69	0.63	0.67
		STI direct	0.96	0.69	0.67	0.79	0.68	0.69	0.67	0.69	0.64	0.67
		RASTI	0.93	0.71	0.68	0.70	0.67	0.71	0.68	0.72	0.64	0.71
		RASTI direct	0.99	0.73	0.68	0.79	0.70	0.72	0.70	0.72	0.65	0.71
Measured	omnidirectional	STI	0.82	0.70	0.69	0.75	0.69	0.67	0.67	0.69	0.64	0.64
		$RASTI_{MLS}$[a]	0.82	0.72	0.74	0.75	0.70	0.69	0.70	0.69	0.64	0.64
	directional	STI	0.89	0.72	0.70	0.78	0.73	0.72	0.71	0.71	0.65	0.67
		$RASTI_{MLS}$	0.91	0.73	0.72	0.80	0.75	0.74	0.74	0.72	0.66	0.70
		$RASTI_{mod}$[a]	0.82	0.70	0.67	0.78	0.71	0.70	0.70	0.69	0.67	0.67

[a]The values presented in Fig. 11 are marked with an a.

ratory rooms. The differences were largest with low speech-to-noise ratios, and, in the worst case, the $RASTI_{mod}$ was 0.13 higher than the $RASTI_{MLS}$. The different directivity of the loudspeakers is the most probable reason for this. This can be seen from the high differences in laboratory measurement 6. In this measurement, there was a lot of absorption material in the reverberation room, although the ceiling and the upper part of the walls were still reflective. The loudspeaker used in the modulated test signal method directs sound upwards more than the one used with the MLS method, and so the $RASTI_{mod}$ values were higher than the $RASTI_{MLS}$ values. At the highest speech-to-noise ratio (>20 dB), which was measured in the room background noise without the masking noise, the speech-to-noise ratio was high enough not to deteriorate speech intelligibility, and so the differences were the smallest.

In the auditorium, the differences between $RASTI_{mod}$ and $RASTI_{MLS}$ were again smaller with the lower background noise than with the higher background noise (Fig. 6). The effect of the randomness of the modulated test signal was not significant, because the measured RASTI values were between 0.67 and 0.82. In the auditorium, the estimated effect of speech spectrum differences was the same as in the laboratory, i.e., small compared with the effect of the directivities of the loudspeakers. The modulated test signal transmitter is supposed to have directional characteristics close to those of a human speaker. The values obtained with the omnidirectional loudspeaker were mostly closer to the values measured with the modulated test signal transmitter than those obtained with the directional loudspeaker. However, the differences between $RASTI_{mod}$ and $RASTI_{MLS}$ were large in the locations far from the normal axis of the loudspeaker measured with both loudspeakers.

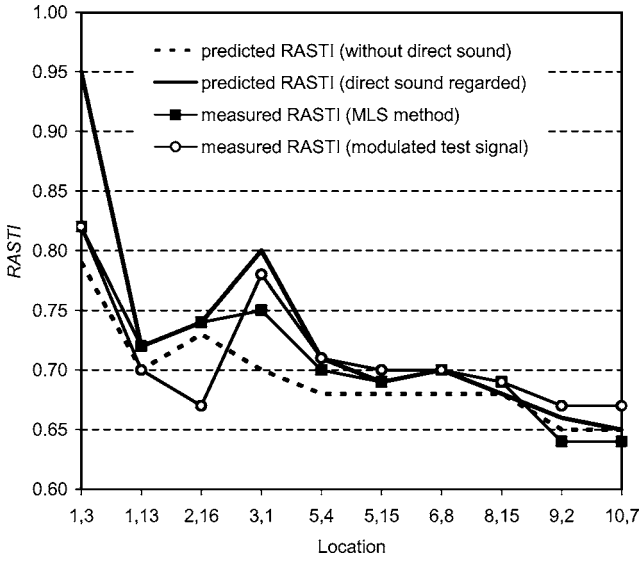


FIG. 11. The predicted and the measured RASTI values in the auditorium. An omnidirectional loudspeaker was used in the MLS method.

The standard deviation of repeated measurements with the modulated test signal was around 0.03, which is acceptable. The background noise conditions were, however, very stable in both the laboratory and the auditorium. This is not necessarily the case in general. Fluctuations in the background noise can result from, e.g., the ventilation system (usually from 0.5 to 2 Hz) or speech carried from other spaces.

In the MLS method, the $RASTI_{MLS}$ is calculated using the same 14 modulation frequencies which are used in the STI calculation. The modulated test signal method uses the nine RASTI modulation frequencies given by the IEC 60268-16 standard. If the MTF decreases smoothly, the RASTI values calculated with the nine modulation frequencies can differ about 0.01 compared with the RASTI values calculated with the 14 modulation frequencies. This estimation is done using Eqs. (4), (6), and (8) and the measured EDTs of the laboratory rooms and the auditorium. If there are very distinct peaks or troughs in the MTF, the selection of modulation frequencies can have a greater effect. More accurate results can naturally be acquired with a larger number of modulation frequencies, as the irregularities in the MTF are then better detected.

With the loudspeakers used in this study, the differences between these two RASTI measuring methods were rather large, especially with low speech-to-noise ratios and with

configurations where the speaker was oriented to the side of the listener. The MLS method was somewhat more laborious, as the speech-to-noise ratio had to be measured separately. The modulated test signal method applies for quick screening purposes as it takes into account the reverberation and background noise simultaneously, but it is more sensitive to fluctuating changes in background noise. However, the fluctuating noise was not a problem in this study.

C. Comparison between the STI and SII

In the laboratory, the differences were, as expected, small between the STI and SII because the speech-to-noise ratio was the same in every frequency band so that the uneven reverberation across the frequency bands was merely the only factor causing the deviations (Fig. 7). In the auditorium, the STI values were always the same or smaller than the SII values (Fig. 8). This was a result of the high reverberation time and low speech-to-noise ratio in the 125-Hz frequency band, which is excluded from the SII calculation. The ANSI S3.5-1997 speech spectrum was used with both the STI and SII for simplicity of measurements and easier comparison. In the auditorium, the STI values would have been at most 0.03 lower than the present values, if the IEC 60268-16 male speech spectrum had been used instead of the ANSI standard speech spectrum. The error is of the same order in the laboratory.

The default band importance functions, which apply for average speech, were chosen for SII calculation. The differences between the STI and SII could have been studied also using the band importance functions for nonsense equally balanced syllables or PB words, but the default values were chosen as they are the most commonly used, and the general difference was of interest.

The SII values were determined using the same modulation frequencies as in the STI calculation but the error is likely to be small. As discussed above, the effect of the selection of modulation frequencies is small, as long as the MTF decreases smoothly. In these laboratory and auditorium conditions, the SII values calculated with the frequencies given by the ANSI standard would be around 0.01 lower than the SII values calculated with the frequencies used in STI calculation, when smooth MTF is assumed.

The ANSI S3.5-1997 standard allows the determination of SII values in low reverberation conditions simply using the measured speech and noise levels. However, no reference values for "low reverberation" are given. In Fig. 12, the SII values as a function of the speech-to-noise ratio, measured in the laboratory rooms with the MLS method, are presented. As the reverberation time decreased, the SII curve approached the theoretical straight line through points (−15 dB, 0) and (+15 dB, 1) as expected according to Eq. (9), when the only affecting disturbance is background noise. For comparison, the SII values were also calculated using only the measured speech and noise levels in the laboratory. These are presented in Fig. 12 with a dashed line. With this simple method, the SII values were essentially the same in every measurement series, because the speech-to-noise ratios were kept the same in every measurement series. The SII values

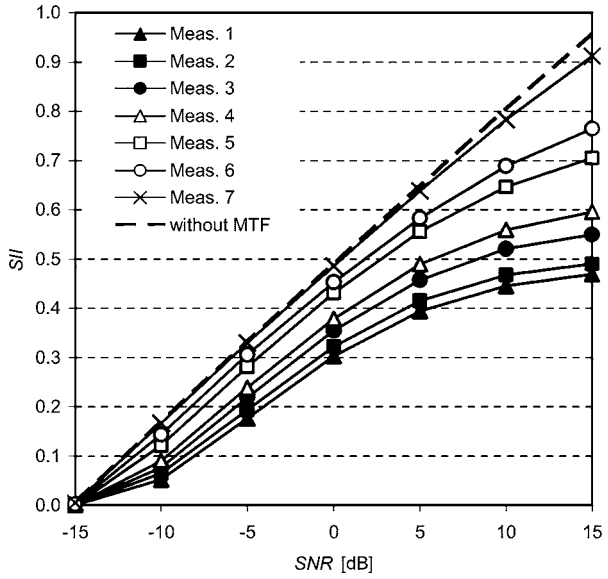


FIG. 12. Measured SII values in the laboratory rooms.

acquired in this simple way correlated sufficiently well with the SII values acquired using the MTF method only in measurement 7, which was conducted in the semi-anechoic room where reverberation was minimal. In real rooms, the reverberation is hardly ever so low. Thus, it is always better to use the MTF method for determination of speech and noise levels in rooms when measuring SII.

D. Comparison between measured and predicted (RA)STI

In Fig. 10, the $RASTI_{mod}$ values are presented using only the 500-Hz frequency band values to illustrate the differences between the measured and predicted RASTI values as a function of the speech-to-noise ratio. The speech-to-noise ratio was not strictly the same in the 2000-Hz as in the 500-Hz frequency band when using the modulated test signal method. According to Eq. (7), the RASTI value should be zero when the speech-to-noise-ratio is −15 dB or less. With the lowest speech-to-noise ratios, the measured RASTI values were clearly higher than the predicted ones. The random properties of the modulated test signal cause that the extreme values 0 and 1 of RASTI are rarely attained. With lower early decay times, predicted RASTI values were close to the measured ones (except for the very low speech-to-noise ratios) but with the highest early decay time, 3.1 s, there was an approximately 0.04 bias in the predicted values, probably because the EDT values used in the prediction were measured with a different loudspeaker from the one used in measuring RASTI.

The room reverberation in the auditorium was not ideally smooth at every frequency band, as can be seen from the 125 and 500-Hz frequency bands in Fig. 9 (and Table II). In the MTF, some peaks or troughs could be seen, which may result from, e.g., echoes. At higher frequency bands, the room reverberation was smoother, like in the 8000-Hz band. Regardless of the nonideal MTF behavior, the predicted STI and RASTI values corresponded quite well to the measured ones in most cases also in the auditorium (Table III). The

absolute deviations between measured and predicted STI and RASTI values were 0.00–0.05 (EDTs obtained with the omnidirectional loudspeaker) and between 0.00 and 0.10 (EDTs obtained with the directional loudspeaker) except for the locations where the receiver was very close to the sound source. In the closest location (1, 3) the largest error was 0.13. In most locations the deviations were smaller when using the EDT values obtained with the omnidirectional loudspeaker and if the direct sound was taken into account.

The prediction was more difficult near the speaker and also far from the axis of the loudspeaker because of the inaccurate directivity assumptions. Elsewhere, the accuracy was good, because the reverberation was still smooth enough in the higher frequency bands and the prediction was made using the measured EDT values and the measured speech and noise levels. If the EDT values and speech and noise levels were also predicted, the accuracy would suffer.

Using the mathematical sensitivity analysis, the RASTI and STI error can be estimated. The absolute error depends on the EDT value, the speech-to-noise ratio, and their absolute errors. If the estimated EDT value absolute error is 0.1 s and the speech-to-noise ratio absolute error 3 dB, the RASTI/STI absolute error is at most 0.11 if the reverberation time is longer than 0.3 s. But if the EDT value error grows to 0.2, the maximum RASTI/STI absolute error grows to 0.15. With very short reverberation times, the error can be very large, but the error decreases as the reverberation time increases.[1]

VI. CONCLUSIONS

Although the RASTI was accurate enough compared with STI values in the study, its feasibility should always be checked, i.e., whether reverberation and speech-to-noise ratios are smooth enough in the room under study.

The selection of the measurement loudspeaker has a great effect on the accuracy of the results of different RASTI methods. The modulated test signal method and the MLS method gave sufficiently close results in the same reverberation conditions, if the speech-to-noise ratio was sufficiently high. The discrepancy between the directivities of the loudspeakers can cause large differences in situations where the speech-to-noise ratio is low and especially when the orientation of the speaker to the listener is not face to face.

STI and SII values are almost the same when the MTF method is used in both methods. The room reverberation is hardly ever low enough for the reverberation not to affect speech intelligibility, so the MTF method should always be used in SII calculation to account for the effect of reverberation, when measuring in rooms and without any speech amplification. The SII and STI are so similar, when used for room acoustical purposes, that the need for many different quantities is questionable.

The STI and RASTI can be predicted rather accurately, if the EDT and the speech-to-noise-values are accurately known at the location of the listener, even though the sound field may not be ideally diffuse.

ACKNOWLEDGMENTS

The authors would like to thank David Oliva Elorza for his help in the measurements and Jarkko Hakala, Riikka Helenius, and Jukka Keränen for technical and consulting assistance during this work. This study was a part of the national research programme Productive Office 2005 (Tuottava Toimisto 2005) funded by Tekes (National Technology Agency), the Finnish Work Environment Fund, and several participating companies.

[1] V. Hongisto, J. Keränen, and P. Larm, "Simple model for the acoustical design of open-plan offices," Acust. Acta Acust. 90, 481–495 (2003).
[2] V. Hongisto, "A model predicting the effect of unwanted speech of varying intelligibility on work performance," Indoor Air (in press).
[3] *ANSI S3.5-1997*, "Methods for calculation of the speech intelligibility index" (American National Standards Institute, New York, 1997).
[4] *IEC 60268-16* 3rd ed., "Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index" (International Electrotechnical Commission, Geneva, Switzerland, 2003).
[5] K. D. Kryter, "Methods for the Calculation and Use of the Articulation Index," J. Acoust. Soc. Am. **34**, 1689–1697 (1962).
[6] *ANSI S3.5-1969*, "Methods for the calculation of the articulation index" (American National Standards Institute, New York, 1969).
[7] N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," J. Acoust. Soc. Am. **19**, 90–119 (1947).
[8] H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326 (1980).
[9] T. Houtgast and H. J. M. Steeneken, "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077 (1985).
[10] H. J. M. Steeneken and T. Houtgast, "Mutual dependence of the octave-band weights in predicting speech intelligibility," Speech Commun. **28**, 109–123 (1999).
[11] "Past, present and future of the Speech Transmission Index," *Proceedings of the International Symposium on STI*, TNO Human Factors, Soesterberg, The Netherlands (2002).
[12] T. Houtgast and H. J. M. Steeneken, "A multi-language evaluation of the RASTI-method for estimating speech intelligibility in auditoria," Acustica **54**, 185–199 (1984).
[13] E. Sala and V. Viljanen, "Improvement of acoustic conditions for speech communication in classrooms," Appl. Acoust. **45**, 81–91 (1995).
[14] M. Hodgson, "Case-study evaluations of the acoustical designs of renovated university classrooms," Appl. Acoust. **65**, 69–89 (2004).
[15] A. P. O. Carvalho, "Relations between rapid speech transmission index (RASTI) and other acoustical measures in churches," Appl. Acoust. **58**, 33–49 (1999).
[16] V. Desarnaulds, "De l'acoustique des églises en Suisse. Une approche pluridisciplinaire," École Polytechnique de Lausanne, Switzerland (2002).
[17] A. Carvalho and C. Monteiro, "Acoustics of courtrooms in Portugal," *Proceedings of the Noise-Con 2003*, Cleveland, OH (2003).
[18] A. Carvalho and A. Carvalho, "Acoustic characterization of historic cloisters in Portugal," *Proceedings of the Noise-Con 2004*, Baltimore, MD (2004).
[19] J. Kang, "Modelling of train noise in underground stations," J. Sound Vib. **195**(2), 241–255 (1996)
[20] S. J. van Wijngaarden and J. A. Verhave, "Measurement and prediction of speech intelligibility in traffic tunnels using the STI," in *Past, Present and Future of the Speech Transmission Index, Proceedings of the International Symposium on STI, TNO Human Factors*, Soesterberg, The Netherlands (2002), pp. 111–116.
[21] P. Larm, J. Keränen, R. Helenius, J. Hakala, and V. Hongisto, "Acoustics in open-plan offices–A laboratory study," *Proceedings of the Baltic-Nordic Acoustic Meeting, 8–10 June*, Mariehamn, Finland CD-ROM (2004).
[22] A. Tisseyre, A. Moulinier, and Y. Rouard, "Intelligibility in various rooms: Comparing its assesment by (RA)STI measurement with a direct measurement procedure," Appl. Acoust. **53**(1), 179–191 (1998).
[23] T. Houtgast, H. J. M. Steeneken, and R. Plomp, "Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics," Acustica **46**, 60–72 (1980).
[24] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Science, New York, 1991).
[25] Brüel and Kjær, "Instruction Manual, Speech transmission meter type 3361" (1986), p. 68.

Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure[a)]

Gisela Jia[b)]
Department of Psychology, Lehman College, City University of New York, 250 Bedford Park Blvd. West, Bronx, New York 10468

Winifred Strange
Graduate Center, City University of New York, 365 5th Avenue, New York, New York 10016

Yanhong Wu[c)]
Department of Psychology, Peking University, Beijing, China, 100871

Julissa Collado
Department of Psychology, Lehman College, City University of New York, 250 Bedford Park Blvd. West, Bronx, New York 10468

Qi Guan
Department of Psychology, Peking University, Beijing, China, 100871

(Received 5 May 2005; revised 18 November 2005; accepted 22 November 2005)

In this study we assessed age-related differences in the perception and production of American English (AE) vowels by native Mandarin speakers as a function of the amount of exposure to the target language. Participants included three groups of native Mandarin speakers: 87 children, adolescents and young adults living in China, 77 recent arrivals who had lived in the U.S. for two years or less, and 54 past arrivals who had lived in the U.S. between three and five years. The latter two groups arrived in the U.S. between the ages of 7 and 44 years. Discrimination of six AE vowel pairs /i-ɪ/, /i-eɪ/, /ɛ-æ/, /æ-ɑ/, /ɑ-ʌ/, and /u-ɑ/ was assessed with a categorial AXB task. Production of the eight vowels /i, ɪ, eɪ, ɛ, æ, ʌ, ɑ, u/ was assessed with an immediate imitation task. Age-related differences in performance accuracy changed from an older-learner advantage among participants in China, to no age differences among recent arrivals, and to a younger-learner advantage among past arrivals. Performance on individual vowels and vowel contrasts indicated the influence of the Mandarin phonetic/phonological system. These findings support a combined environmental and L1 interference/transfer theory as an explanation of the long-term younger-learner advantage in mastering L2 phonology. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151806]

PACS number(s): 43.71.Hw, 43.71.Ft, 43.70.Ep [ALF] Pages: 1118–1130

I. INTRODUCTION

In cross-language developmental studies of non-native speech learning, two primary research goals are to accurately document and explain developmental changes in the ability to learn new speech sounds. To address these goals, the current study investigated how age-related differences vary along one important dimension of learning, the amount of exposure to the target sounds.

Past research on age-related differences in non-native speech learning can be classified into two main categories: laboratory studies and immersion studies. In laboratory studies, participants live in their native country and have no immersion experience[1] with the target language. Participants are exposed to the target speech sounds of a foreign language only in the study setting, usually a research laboratory. This approach allows the assessment of age differences at the initial encounter with the new sounds, and offers good control over the amount of exposure to these sounds. Laboratory studies, which have focused primarily on production, have yielded inconsistent findings. Findings from some studies support the notion of "the younger the better." When imitating Spanish words, monolingual English-speaking 7 year olds were slightly but significantly more accurate than young adults (Cochrane and Sachs, 1979). Similar findings were obtained among a group of native English-speaking 5–15 year olds when imitating French and Armenian words and phrases (Tatha, Wood, and Loewenthal, 1981a). In contrast, findings from other studies support the notion of "the older

[a)]A portion of this work was published in "Age differences in perceptual sensitivity to new speech sounds: The younger the better?" Proceedings of the 29th Boston University Conference on Language Development, Boston, November, 2004, and was presented in "Age differences in the perception and production of American English vowels by native Mandarin speakers." Poster presentation at the 1st Acoustical Society of America Workshop on Second Language Speech Learning, Vancouver, Canada, May, 2005.

[b)]Corresponding author. Electronic mail: giselajia@yahoo.com
[c)]Corresponding author. Electronic mail: wuyh@pku.edu.cn

the better." The accuracy in imitating Dutch words increased linearly with age among native English speakers ranging in age from 7 years to young adulthood (Snow and Hoefnagel-Höhle, 1977). The ability to imitate French words and discriminate French-sound pairs also increased with age among English-speaking first to ninth graders (Politzer and Weiss, 1969).

Taking advantage of the immigration phenomenon, immersion studies have examined age-related differences exhibited by immigrants as they acquire a new language in the second language (L2)-speaking country. These studies investigate the relation between age of exposure to L2, usually indexed by age of arrival (AoA) in the L2-speaking country, and learners' L2 speech perception and production abilities. Immersion studies can be conducted at different points in time along a wide spectrum of length of L2 immersion. Long-term immersion studies include L2 learners who have resided in the L2-speaking country for many years when their L2 proficiency supposedly has reached relative stability following massive L2 exposure. Short-term immersion studies include L2 learners at a more recent stage of L2 immersion. There is no clear cut division between short and long terms, as some studies adopt a five-year (e.g., Jia, Aaronson, and Wu, 2002) and others a ten-year criterion (e.g., Flege, Munro, and MacKay, 1995a).

Findings from long-term immersion studies have consistently shown that, when the length of residence in the L2 country being equal, younger arrivals obtain better L2 speech perception and production skills than older arrivals. The benefit of early arrival existed for the overall degree of perceived foreign accent by English L2 learners speaking various native languages (Asher and Garcia, 1969; Flege *et al.*, 1995a; Oyama, 1976; Yeni-Komshian, Flege, and Liu, 2000), and for the accuracy in the perception and production of American English (AE) vowels and consonants (Flege, MacKay, and Meador, 1999; Flege, Munro, and MacKay, 1995b; MacKay, Flege, Piske, and Schirru, 2001; Munro, Flege, and MacKay, 1996) by native Italian speakers. Such age-related differences in production exist at an even earlier point of L2 immersion, i.e., after about two to three years of L2 immersion (Fathman, 1975; Tatha, Wood, and Loewenthal, 1981b).

Different from long-term attainment and laboratory studies that examine the performance at the one-time point, longitudinal immersion studies track performance over time. Snow and Hoefnagel-Höhle (1977) studied native English speakers living in Holland ranging in age from three years to adulthood. At three time points with a four to five month interval, participants distinguished Dutch minimal pairs, as well as imitated and spontaneously produced Dutch words. Although there were no significant age-related differences for perception at any point, age-related differences in production changed with increasing immersion experience. At the first testing session, older children and adults did significantly better than younger children in pronouncing many vowels and consonants. At the second session, age differences in pronouncing most of the segments disappeared. At the third session, age differences became reversed, with younger children outperforming older children and adults.

More recently, Flege *et al.* (in press) studied 155 native Korean-speakers living in the U.S. and Canada. The child arrivals (AoA between 6–12 years) and adult arrivals (AoA between 21–35 years) were tested after 3–4 years and then at 5 years of residence in these countries. The adult arrivals were judged to speak English with a significantly stronger foreign accent than the child arrivals. In a subgroup of these participants ($n=108$), the ability to discriminate and imitate English vowels was examined. Child arrivals outperformed adult arrivals after both 3 and 5 years of residence on both perception and imitation (Tsukada *et al.*, 2005). Similar to Snow and Hoefnagel-Höhle (1977), these two studies demonstrated a period of younger-learner advantage. Different from Snow and Hoefnagel-Höhle, these two studies did not observe a period of older-learner advantage. This is likely due to the fact that the time 1 of the Flege *et al.* and Tsukada *et al.* studies was already after 2 years of L2 immersion, when the adult advantage could well have already disappeared. Indeed, in a longitudinal study of native Japanese speakers' perception and production of English consonants /l/, /r/, and /w/, adult arrivals performed significantly better than child arrivals after six months of L2 immersion. However, after a year, the trend was reversed (Aoyama *et al.*, 2004).

In sum, the few existing longitudinal studies suggest that age-related differences may change with increasing L2 exposure. The extent to which a study can demonstrate the crossover pattern depends on the time point(s) selected for the study. In the beginning of L2 immersion, older learners may have an advantage in discriminating and producing nonnative speech sounds. This advantage may persist in early years of L2 immersion, but disappear and become reversed later. The exact timetable of this crossover pattern may depend on which aspects of phonological skills are assessed and the difficulty level of the assessment tasks.

Three general theoretical accounts have been put forth to explain the robust early learner advantage found in long-term immersion studies: the Critical/Sensitive Period account, the L1 Transfer/Interference account, and the Environmental account. The Critical/Sensitive Period account postulates a genetically guided maturation of a domain specific language learning mechanism as the reason for the declining abilities to learn new speech sounds. That is, there is a "limit to the programming of new phonological rules at the cortical level" (Cochrane, 1980; pp. 332–333) due to factors outside the learning process (e.g., Patkowski, 1990; Scovel, 2000).

In contrast, the L1 Transfer/Interference account posits that "new processing activities are hampered by the strength of heavily utilized patterns or processing routines" (Cochrane, 1980; pp. 332–333) used in native language processing. This account has been represented by three models of cross-language speech perception, the Perceptual Assimilation Model (PAM) (Best, 1995), the Native Language Magnet Model (NLM) (Kuhl, 2000; Werker and Tees, 1999), and the Speech Learning Model (SLM) (Flege, 1995). The three models, with some nonoverlapping foci, converge on an emphasis on the role of prior L1 learning and use in L2 speech perception and production. According to PAM, non-native segments "tend to be perceived according to their similarities

to, and discrepancies with, the native segmental constellations that are in the closest proximity to them in native phonological space." (Best, 1995, p. 193). As will be referred to in detail later, this model makes specific predictions about the relative difficulty with which particular non-native segments are perceived or produced, based on their relation to the native phonological system. NLM delineates the details as to how, as early as in their first year of life, infants form a complex perceptual network through which new speech sounds are perceived, or "filtered" (Kuhl, 2000; Werker and Tees, 1999). SLM focuses in part on explaining age-related differences in learning new speech sounds. According to SLM, the greater difficulties experienced by older individuals arise from the increasingly strong influence of L1 (Flege, 1995). More specifically, with increasing age, L1 phonetic categories exert stronger assimilation power on non-native speech sounds, making the establishment of new speech categories more difficult (Baker, Trofimovich, Mack, and Flege, 2002; Flege, 2003).

The Environmental account (Jia and Aaronson, 2003; Snow, 1983) has been developed to explain the younger-learner advantage in various aspects of L2 proficiency found in long-term attainment studies. According to this account, in the immigration setting, L2 learners of various ages are inherently at different levels of cognitive, social, and cultural maturation. Such variations expose early arrivals to a significantly richer L2 environment than late arrivals, and such environmental differences accumulate and lead to L2 proficiency differences.

The validity of these three accounts relies heavily on a more accurate description of age-related differences in L2 speech learning. If the age difference crossover pattern discussed earlier proves robust, all three theories need to address these related questions. Why are early arrivals better in the long run? Why do late arrivals initially have an advantage? Why does it take time for early arrivals to catch up with and eventually surpass late arrivals? Most previous studies have focused on a limited period of L2 exposure. This has prevented firm conclusions about the interaction between age-related differences and the amount of L2 exposure. Although we can summarize trends from different studies, their sampling of different language populations and use of different methods and designs weaken the conclusion. Therefore, to shed light on the validity of the theoretical accounts, it is important to conduct further research to capture the dynamic changes of age-related differences in L2 speech learning at different points of L2 immersion with diverse populations.

To examine the interaction of age and amount of exposure to the target language, the current study included three participant groups with different amounts of exposure to native-sounding AE. The first was a group of native Mandarin speakers (chronological age at time of study 7 to 20 years) living in the People's Republic of China (hereinafter referred to as China) with no English immersion experiences. Age differences found in this group are not confounded by age-related language environment differences existing in immigrant populations (Jia and Aaronson, 2003). Further, from the perspective of immigration, these participants could be a potential group of immigrants on the first day of their arrival in the L2-speaking country. Their chronological age at the time of the study matched the AoA for immigrant L2 learners. The study also included two groups of native Mandarin speakers who immigrated to the U.S. at various ages: recent arrivals with a length of U.S. residence of 2 years or less, and past arrivals with a length of U.S. residence between 3 and 5 years. Based on the previous research findings, we anticipated a crossover pattern of age differences among these three groups.

In the current study we focused on the perception and production of L2 vowels. Previous research has shown that many L2 learners have persistent problems with L2 vowels (e.g., Bohn and Flege, 1992; Strange *et al.*, 1998). Vowel production properties (spectral and temporal) have been found to contribute to the intelligibility of L2 English produced by native Mandarin speakers (Rogers, 1997) and Japanese speakers (Kewley-Port, Akahane-Yamada, and Aikawa, 1996). The vowels selected in the current study also bear different relations to the L1 vowel space, and thus allowed us to test theories (e.g., PAM, SLM) that predict specific types of L1 influences on L2. The details of these vowels will be presented in Sec. II.

Non-native speech perception and production is influenced by speaker and contextual variability in the target language. In order to pinpoint the source of vowel perception and production problems, the current study minimized these variations by conducting the tasks under constrained processing conditions. The target vowels were read by only one speaker in one consonantal context to form nonsense words for discrimination and immediate imitation. To further reduce the processing demands, perception and production procedures that minimize the memory load were used. If participants have problems discriminating and imitating vowels in this highly structured, predictable, and controlled context, they would experience greater difficulties with these vowels when placed in longer speech units during speech communication. Thus, in the study we examined the initial sources of difficulties experienced by participants, as well as whether and in what ways age-related differences exist at this processing level.

II. METHOD

A. Participants

1. Native Mandarin speakers in China

Participants in China were 91 native Mandarin-speaking children, adolescents, and young adults from 7 to 20 years of age. There were 46 females and 45 males, fairly evenly distributed across the age span with about six participants in each age group. Born and raised in Beijing, all participants spoke Peking Mandarin. They were recruited from average quality elementary, middle, and high schools, and colleges in Beijing. Participants were reported by teachers to have average level academic performance. To minimize the participants' exposure to native-sounding English, participants were from schools with no native-English speaking teachers and no extensive training in English listening. No participants had received private English lessons with native En-

TABLE I. Demographic information for native Mandarin speakers in the U.S. (past arrivals and recent arrivals).

Demographic variables	Past arrivals (n=54) Mean	SD	Range	Recent arrivals (n=77) Mean	SD	Range
Age (year)	24.4	8.0	11–44	20.5	8.7	8–46
Age of immersion (year)	20.7	8.0	7–40	19.2	8.7	7–44
Age of instruction (year)	11.4	2.7	6–24	11.4	2.1	6–16
U.S. stay (year)	3.7	0.8	3–5	1.3	0.7	0–2
U.S. education (year)	2.3	1.5	0–6	0.9	0.9	0–2

glish speakers, and few had attended supplementary English classes outside of school. The number of years of English language instruction ranged from 0 to 11 years (M=4.41 year; SD=2.81), mostly beginning in the fourth (36.4%), third (22.70%), and first (28.40%) grades.

2. Native Mandarin speakers in the US

Participants in the U.S. were 131 native Mandarin speakers who immigrated to New York City (NYC) between 7 and 44 years of age and had lived in the U.S. for fewer than 5 years. They were divided into two subgroups according to their length of U.S. residence: 54 past arrivals who had lived in the U.S. for between 3 and 5 years, and 77 recent arrivals who had lived in the U.S. for two years or less (Table I). These two groups did not differ significantly in their age, AoA, and age of onset of English instruction. They were set apart by years of residence in the U.S., and consequently, years of education in the U.S. Participants were recruited from the Chinese communities in NYC through an advertisement in a Chinese newspaper. The majority (n=99) spoke Mandarin (75.57%) as their native dialect, 14 (10.69%) spoke Min dialect, 13 (9.92%) spoke Wu dialect, and 5 (3.82%) spoke Cantonese. All non-native Mandarin speakers were exposed to Mandarin from birth, and all started speaking Mandarin regularly in school before 9 years of age. Similar to participants in China, their exposure to native-sounding English before their arrival in the U.S. was minimal.

No hearing screening was conducted for participants in China or the U.S. However, all participants reported having normal hearing in a background questionnaire described later.

B. Stimulus materials

The AE vowel inventory can be described as including 11 nonrhotic monophthongal vowels differing in height (5 levels: high, mid-high, mid, mid-low, low) and position (front versus back). The front vowels are /i, ɪ, e, ɛ, æ/ and the back vowels are /u, ʊ, o, ɔ, ʌ, ɑ/. The mid vowels /e, o/ are usually phonetically realized as diphthongal [$e^{ɪ}$, $o^{ʊ}$] in stressed syllables, mid-low /ʌ/ is unrounded and centralized relative to mid-low rounded /ɔ/, and other vowels show some diphthongization in some dialects (Peterson and Barrey, 1952; Hillenbrand, Getty, Clark, and Wheeler, 1995). The duration of AE vowels also varies phonetically, with the four short vowels [ɪ, ɛ, ʊ, ʌ] alternating with the seven long vowels [iː, eː, æː, uː, oː, ɔː, ɑː] (Peterson and Lehiste, 1960). Mandarin has a smaller vowel inventory than AE. The description of the Mandarin vowel inventory has been controversial, due to different classification criteria and methods of analysis, but researchers generally adopt a six-vowel system (Howie, 1976; Lin, 1989; Wan, 1999): three high vowels—front unrounded /i/, front rounded /y/, and back rounded /u/; two mid-vowels—central /ə/ and back /ɤ/; and one low vowel /a/. Allophonic variations of /i/ include high and mid-high variants [i, ɪ]; central /ə/ varies allophonically from mid-central to mid-front [ə, e]; low /a/ varies allophonically from central to back variants [a, ɑ]; mid-back /ɤ/ includes both unrounded and rounded allophones [ɤ, o]. Mandarin vowels appear in open syllables /CV, V, VV/, and /Vn/ and /VVn/ syllables. Vowel duration does not distinguish vowels in Mandarin. Stimulus materials for the current study included eight AE vowels /i, ɪ, $e^{ɪ}$, ɛ, æ, ʌ, ɑ, u/, of which /i, u/ have phonetically similar counterparts in Mandarin. The other six vary in their relation to Mandarin vowels. The mid-low and low front vowels /ɛ, æ/ and the mid-low back /ʌ/ have no phonetically similar counterparts, even when allophonic variation is taken into consideration. The AE vowels /ɪ, $e^{ɪ}$, ɑ/ are phonetically similar to contextual variants of Mandarin phonemes.

For the perception task, six contrasts were formed with these eight vowels, including /i-ɪ/, /i-$e^{ɪ}$/, /ɛ-æ/, /æ-ɑ/, /ɑ-ʌ/, and /u-ɑ/. The vowel contrasts were selected to present a wide range of difficulty for native Mandarin speakers, according to data from the few studies of native Mandarin speakers (Rogers, 1997; Flege, Bohn, and Jang, 1997), and the predictions of the relevant theoretical models, such as PAM (Best, 1995), and SLM (Flege, 1995). The vowels in the /ɛ-æ/ and /ɑ-ʌ/ pairs are close in articulatory and acoustic vowel space, and neither occurs in Mandarin (though /ɑ/ occurs in Mandarin as an allophonic variant of /a/). Vowels in the /i-ɪ/ pair are close in vowel space, but /i/ occurs in Mandarin. The pairs /i-$e^{ɪ}$/ and /æ-ɑ/ are more distant in vowel space than the preceding pairs, and /i/ occurs in Mandarin. Finally, the /u-ɑ/ vowels are distant in vowel space and have distinctive counterparts in Mandarin. In terms of Best's PAM, we speculated that the first two pairs fall into a Single Category Assimilation pattern, the next three pairs into a Category Goodness pattern, or for /i-$e^{ɪ}$/, possibly a Two-Category pattern, and the final contrast is a clear Two-Category pattern. PAM predicts the order of difficulty for these contrasts as (from most to least difficult): /ɛ-æ/, /ɑ-ʌ/, /i-ɪ/, /æ-ɑ/, /i-$e^{ɪ}$/, and /u-ɑ/.

The selected vowels were situated in /dV-pə/ disyllables spoken in citation form. The use of nonsense disyllables rather than real words minimized lexical effects (the effects of previous learning on task performance). Importantly, the /dV-pə/ syllabic structure also conforms to both Mandarin and English phonotactic constraints. Multiple tokens of each vowel stimulus were used to approximate the inherent generalization processes required during speech learning. One female monolingual AE speaker with no noticeable regional accent recorded the stimuli /dæpə/, /dɛpə/, /dʌpə/, /dɑpə/, /dɪpə/, /dipə/, /de^{ɪ}pə/, and /dupə/ in this sequence ten times. The stimuli were recorded directly as digitized sound files (22.05 kHz). All disyllables were read in a flat intonation pattern, similar to the level lexical tone in Man-

darin. The first two sets were used for practice, and the other eight sets were analyzed for five acoustic parameters of the target vowel (VOT, length, pitch, and Fl and F2 values), and two acoustic parameters of the nontarget vowel /ə/ (VOT and length). For each target vowel, three tokens were selected out of the eight tokens (see the Appendix). In order for a token to be selected, the target vowel had to have a minimum of four acoustic parameter values within the 95% confidence interval of the mean, and the nontarget vowel had to have as many as possible (ranging from 0–2) acoustic parameter values within the 95% confidence interval.

C. Design and procedure

1. Perception

Perception accuracy was assessed using a categorial (name identity) AXB discrimination task. This task was chosen among several discrimination tasks because it avoids the possibility of an age-related criterion shift found in same-different judgment tasks (Beving and Eblen, 1973) and possible difficulties that young children may have in understanding the concepts of "same" and "different." Further, an AXB task poses less memory and processing demands than the other two triplet formats (Oddity, ABX) because the middle target stimulus is next to both comparison stimuli (MacKain, Best, and Strange, 1981).

Each vowel pair was tested with 12 trials, 3 trials for each of the 4 possible position combinations (AAB, ABB, BAA, BBA). This resulted in 72 trials for the whole test (6 pairs $\times$ 4 position combinations $\times$ 3 trials). The 72 trials were presented in 6 blocks of 12 trials. Each vowel pair appeared twice in each block. The order of blocks and trials within each block were randomized across participants. Each of the three selected tokens of a vowel was used the same number of times. Vowel positions were also balanced within and across blocks. The two same vowels in each AXB triplet were always two physically different stimulus tokens. This allowed us to test categorial perception at the minimum level, though not to the full extent as no differences in speakers or consonantal context were included.

A block of 12 trials with five Mandarin vowels /i, y, ɘ, a, u/ designed in exactly the same format was presented before the test to familiarize participants with the task as well as to screen participants. Participants who made three errors or more were allowed to proceed with and complete the entire study, but their data were not included in analyses. According to the above criterion, four participants in China (one 8-year-old, two 9-year-olds, and one 15-year-old) were excluded from data analyses, leaving 87 participants for this group.

The AXB task was conducted using specialized computer software (written by Bruno Tagliaferri) available in the Speech Acoustics and Phonetics Laboratory (SAPL) at the CUNY Graduate Center. Each stimulus triad was preceded by a tone presented 300 ms prior to the first stimulus. After listeners heard the three disyllables (ISI=500 ms), two boxes appeared on the screen. The left one read "1," and the right box read "3." Participants were instructed to click "1" if they decided that the middle disyllable sounded like the first one, and click "3" if the middle one sounded like the third one. Once the participants clicked "1" or "3," the next trial was triggered, with a 1000 ms intertrial interval. The trial and test sessions together took between 10 and 15 min. After each block of 12 trials, participants were offered the choice to take a break, although no participant chose to do so. All participants were tested individually, listening to the stimuli through earphones with volume adjusted to a comfortable level for the individual.

Participants in China were tested in a quiet office in their schools in Beijing, on a 15-in. screen portable PC. Participants in U.S. were tested in a soundproof room in the CUNY laboratory, using a 19-in. screen desktop PC.

2. Production

Prior to the discrimination task, participants imitated each of the eight /dV-pə/ stimuli (/dæpə/, /dɛpə/, /dʌpə/, /dɑpə/, /dɪpə/, /dipə/, /deɪpə/, and /dupə/) three times consecutively, each time immediately after hearing the target disyllable. The production tokens were directly recorded as digitized sound files (22.05 kHz, 16-bit resolution) and then normalized for peak amplitude using Sound Forge. The files were further processed for an identification task by native English speakers. The files were first sliced into separate sound files each with one disyllable. Then, the nontarget vowel in each disyllable was removed by deleting all portions of the signal following the beginning of the /p/ stop closure defined as the cessation of upper formant energy. The aim of the editing was to eliminate the potential distraction of the nontarget vowel from the focus on the target vowel. Finally, each file was duplicated so listeners heard each stimulus twice. The time interval between the repetitions was 1000 ms.

For the purposes of token and response choice selections, a pilot identification task was conducted. Three native English-speaking listeners with IPA knowledge heard all three tokens of each vowel produced by the Mandarin speakers in China. A total of 16 AE monophthongs and diphthongs were used as response choices. Among the three tokens produced for each vowel, the second token elicited the highest agreement rate among the judges, and also yielded the most consistent identification results with both the first and the third token. Therefore, to reduce the amount of testing time, only the second repetition of each vowel was selected for the final task. Further, four of the 16 response choices that were never chosen by any listener were eliminated from the final identification choices.

There were a large number of clipped sound files for the participants in China. To counter their tendency to speak softly during the recording, they were instructed to speak loud, risking some signals being clipped. The productions of 42 participants in China and 127 participants in the U.S. who had at least one good token of each vowel were used. This yielded 168(42+126) $\times$ 8 utterances. These utterances were blocked by speakers, with 8 trials in each block. The productions were divided into four sessions with an approximately equal number of blocks. Each session had similar proportions of tokens produced by speakers from each group (speakers in China, recent arrivals and past arrivals), age (for

speakers in China), AoA (for speakers in the U.S.), and gender. When presented to the listeners, the order of the blocks and trials within a block were all randomized separately for each listener.

The 1344 utterances (168 participants × 8 vowels) were presented to five native speakers of English with a mean age of 39.4 years. Three listeners grew up in NYC and spoke English with the local accent. The other two were raised in Chicago or New Jersey, but both were familiar with New York City accent. All had IPA knowledge but were not experienced phoneticians. All listeners reported normal hearing. They listened to the tokens individually in an IAC acoustic chamber using customized software (written by Bruno Tagliaferri) that controlled stimulus presentation and recorded responses to an Excel data form. They completed two sessions on each of two separate days with a brief break between sessions. Listeners heard the stimuli through headphones at a comfortable level. They were instructed to pay attention to the vowel in the syllable, and identify, among the 12 orthographic labels and IPA symbols ("deep /dip/," "dip /dɪp/," "dape (date) /deᵻp/," "dep (debt)/dɛp/," "dap (dash) /dæp/," "dop (dock) /dɑp/," "dup (duck) /dʌp/," "dawp (dawn) /dɔp/," "dope (doze) /dop/," "doop (food) /dup/)," "dUp (could) /dʊp/," "dype (diaper) /daɪp/"), the one that sounded closest (though maybe not identical) to the token just heard. Before the test, listeners completed five practice blocks of 40 trials (5 speakers × 8 tokens) to familiarize themselves with the task. For the five speakers whose productions were used for the practice blocks, one was a monolingual English speaker who produced the stimuli for the current study, four were native Mandarin speakers (one adult male, one adult female, one child male, and one child female). Their imitation of the nonsense disyllables were elicited in exactly the same condition as the formal participants. Responses to these practice trials were not included in the data analyses. All five native listeners identified all the tokens of the monolingual English speaker correctly. Due to dialect influences, two additional native English listeners each made one or two errors identifying the monolingual tokens. These two listeners did not proceed with the identification study.

3. Background questionnaire

After the production and perception sessions, all participants filled out a background questionnaire. The questionnaires for participants in China and NYC were not identical but had overlapping items. The common items included gender, birth date, birth place, places where participants had lived, and any known hearing and health problems. Participants in China, in addition, listed their current school grade, the grade that English instruction began, and the number of hours of English classes in each week. Participants in NYC provided information about their age when English language instruction began, and their age of arrival in the U.S. Children and adolescents living with their parents rated their mothers' and fathers' English speaking ability along a 1–7 point scale (1 = cannot speak English at all; 7 = speak English as fluently as a native English speaker). They also reported the percentage of time that their father, mother, and siblings spoke to them in English, and the percentage of time that they watched TV and videos in English. All the above items regarding parents' English proficiency and language use were rated for every year that participants were in the US. The average English use in a situation across all the years of U.S. residence was calculated for use in the statistical analyses.

III. RESULTS

The results are organized into three sections: perception, production, and the relation between perception and production. For both perception and production, performance accuracy was compared among the groups and across the vowel pairs (perception) or vowels (production) using mixed Analyses of Variance. Age-related differences were examined within each group using bivariate correlations, and other predictive variables were also investigated using regression analyses. Correlation and regression analyses were chosen over age group analyses because the former treats age as a continuous variable and maximizes its variance. The relation between perception and production was examined at the individual level as indicated by correlations between performance on perception and production, and at the group level by the extent to which the rank order of vowel pair (or vowel) difficulties matched in perception and production. In all results of the Analyses of Variance (ANOVA) reported below, the effect size (ES) is indicated by eta-squared values (η^2).

A. Perception

1. Accuracy across groups and vowel pairs

Perception accuracy was indicated by the percentage of correct responses out of the total 72 trials (for total accuracy), or the 12 trials (for each vowel contrast). Performance accuracy for the total task and for each vowel pair was compared across three participant groups. All three groups performed well above chance level with over 70% accuracy for all contrasts (Table II). A mixed two-way 6 × 3 ANOVA was conducted, with the 6 vowel pairs as the within-subjects variable, and the three participant groups as the between-subjects variable. The results revealed a main effect of group, $F(2,215)=53.18$ ($\eta^2=0.33$), a main effect of pairs, $F(4,862)=101.98$($\eta^2=0.32$; with Greenhouse–Geisser correction of degrees of freedom), and an interaction between group and pairs, $F(8,862)=10.87$ ($\eta^2=0.09$; with Greenhouse–Geisser correction of degrees of freedom) (all $p<0.001$).

The main group effect reflects differential performance across the three groups. Pairwise comparisons (with Bonferroni corrections) indicate that participants in China had significantly lower accuracy than the recent and past arrivals. To further examine the group effect for each vowel pair, separate one-way ANOVA was performed for the performance on individual vowel pairs. There was a significant effect of group for all vowel pairs, including /i-eᵻ/, $F(2,215)=26.97$, /æ-ɑ/, $F(2,215)=11.66$, /ɑ-ʌ/, $F(2,215)=32.46$, /i-ɪ/, $F(2,215)=43.25$, /ɛ-æ/, $F(2,215)=32.77$, and /u-ɑ/, $F(2,215)=8.29$ (all $p<0.001$). Bonferroni post-hoc tests re-

TABLE II. Performance on the six contrasts by native Mandarin speakers in China (monolinguals) ($n=87$), recent arrivals ($n=77$), and past arrivals ($n=54$).

Vowel pairs	Monolinguals ($n=87$) % correct (SD; range)	Recent arrivals ($n=77$) % correct (SD; range)	Past arrivals ($n=54$) % correct (SD; range)
/i-ɪ/	82.6 (16.9; 33.3–100)	97.4 (6.4; 66.7–100)	97.8 (5.4; 66.7–100)
/i-eɪ/	90.2 (14.3; 41.7–100)	99.5 (2.1; 91.7–100)	99.7 (1.6; 91.7–100)
/ɛ-æ/	76.3 (14.2; 41.6–100)	89.4 (12.9; 33.3–100)	91.8 (9.5; 58.3–100)
/æ-ɑ/	88.9 (14.4; 41.7–100)	96.1 (8.7; 50.0–100)	96.6 (7.7; 58.3–100)
/ɑ-ʌ/	71.7 (16.5; 33.3–100)	85.2 (16.5; 25.0–100)	91.4 (9.4; 58.3–100)
/u-ɑ/	97.9 (4.6; 75.0–100)	99.7 (1.6; 91.7–100)	99.7 (1.6; 91.7–100)
Overall	84.6 (10.4; 55.6–98.6)	94.5 (5.5; 69.4–100)	96.2 (3.3; 87.5–100)

vealed that, for all vowel pairs, both the recent and past arrival groups scored significantly higher than the China group, and there were no significant differences between the two immigrant groups, probably due to ceiling effects.

Regarding the main effect of pairs, pairwise comparisons (with Bonferroni corrections) revealed that scores (averaged across the three participant groups) on most pairs of vowel contrasts (except for /æ-ɑ/ and /i-ɪ/) were significantly different (all $p<0.001$). The interaction effect of group and pair indicates that the performance difference in pairs varied among the three groups. To further examine this effect, paired-sample T tests were conducted separately for each participant group to compare the performance on each pair of vowel contrast. For the participants in China, only one difference between vowel contrasts (/i-eɪ/-/æ-ɑ/) were not significant, and all other 14 pairs were significant (all $p<0.01$). For the recent arrivals, two difference scores (/i-eɪ/-/u-ɑ/; /i-ɪ/-/æ-ɑ/) were not significant. For the past arrivals, three difference scores (/i-eɪ/-/u-ɑ/; /i-ɪ/-/æ-ɑ/; /ɛ-æ/-/ɑ-ʌ/) were not significant. In terms of the rank order of performance, the two most difficult pairs (/ɛ-æ/ and /ɑ-ʌ/) and the two easiest pairs (/i-eɪ/ and /u-ɑ/) were the same for all three groups. The difficulty order for two medium-level performance pairs (/i-ɪ/ and /æ-ɑ/) was the opposite for participants in China and the U.S.

2. Age differences

The age variable of interest is the age of L2 exposure. For recent and past arrivals, it was indicated by AoA in the L2-speaking country. For participants in China, it was indicated by chronological age at the time of the study, which coincides with AoA, as they could be regarded as a group of immigrants on their first day of arrival in the U.S. Participants in China showed significant positive correlations between age and performance on the total task ($r=0.51$, $p<0.001$) and on all the individual vowel contrasts ($r=0.39$, $p<0.001$ for /i-ɪ/, $r=0.38$, $p<0.001$ for /i-eɪ/, $r=0.37$, $p<0.001$ for /ɛ-æ/, $r=0.43$, $p<0.001$ for /æ-ɑ/, $r=0.41$, $p<0.001$ for /ɑ-ʌ/, and $r=0.28$, $p<0.01$ for /u-ɑ/), indicating that older participants, in general, achieved a higher level of accuracy (Fig. 1). However, recent arrivals showed no significant correlations between AoA and overall performance or individual vowel pairs. In contrast, past arrivals showed negative correlations between AoA and overall performance ($r=-0.41$, $p<0.01$), and two of the more difficult vowel contrasts ($r=-0.36$, $p<0.01$ for /æ-ɑ/, and $r=-0.40$, $p<0.01$ for /ɑ-ʌ/), a trend opposite that of the participants in China. That is, a younger AoA predicted significantly better performance on the task in general, and on the difficult vowel contrasts.

3. Other predictors

To pinpoint the unique predictive power of AoA, several other potential predictor variables of performance were also examined. For the participants in China with little variance in the English environment, the major variable was length of English instruction. Older participants had significantly more years of English instruction, $r=0.89$, $p<0.001$, and more years of English instruction predicted a better task performance, $r=0.48$, $p<0.001$. A partial correlation analysis was conducted to examine the unique contribution of chronological age (AoA in our definition) and length of English instruction. When the length of English instruction was partialed out, there was still a marginally significant relation between age and the total percentage correct, $r=0.19$, $p=0.07$. When age was partialed out, the correlation between years of English instruction and the total percent correct became nonsignificant ($r=0.07$, $p=0.50$).

The predictor variables for the two immigrant groups included the age that English instruction began, the length of U.S. residence, the length of U.S. education, parents' English speaking abilities, and the percentage of English use in various situations. Bivariate correlations between total accuracy and all of these predictive variables were obtained for each

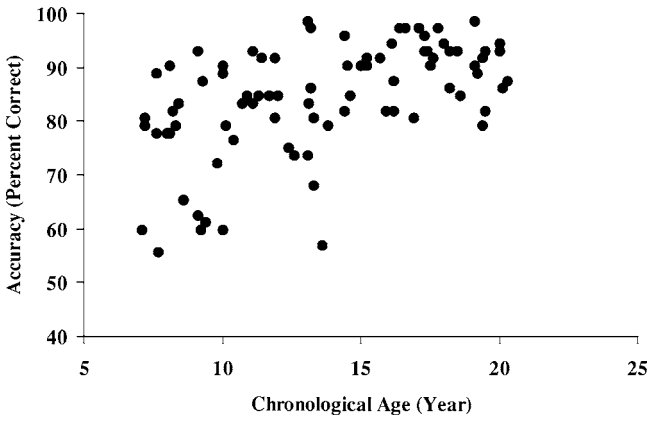


FIG. 1. Scatter plot of age and total accuracy (percentage correct) for native Mandarin speakers in China ($n=87$; $r=0.51$, $p<0.001$).

TABLE III. Production accuracy of all the vowels and the individual vowels by native Mandarin speakers in China (monolinguals) ($n=42$), recent arrivals ($n=76$), and past arrivals ($n=50$).

Vowel pairs	Monolinguals ($n=42$) % correct (SD; range)	Recent arrivals ($n=76$) % correct (SD; range)	Past arrivals ($n=50$) % correct (SD; range)
/i/[a]	88.1 (23.8; 0–100)	88.4 (24.7; 0–100)	89.2 (21.5; 0–100)
/ɪ/	52.4 (42.6; 0–100)	76.3 (35.0; 0–100)	77.2 (35.5; 0–100)
/eɪ/	54.3 (42.0; 0–100)	88.9 (15.5; 40.0–100)	88.4 (18.6; 20–100)
/ɛ/	65.2 (32.8; 0–100)	65.8 (33.3; 0–100)	72.0 (27.1; 0–100)
/æ/	52.4 (38.7; 0–100)	64.7 (37.5; 0–100)	68.8 (33.6; 0–100)
/ɑ/	51.4 (37.0; 0–100)	77.1 (29.5; 0–100)	74.0 (33.9; 0–100)
/ʌ/	54.3 (31.0; 0–100)	44.7 (34.2; 0–100)	48.8 (36.4; 0–100)
/u/[a]	97.1 (9.4; 60–100)	92.6 (16.6; 20–100)	95.2 (17.4; 0–100)
Overall	64.4 (11.1; 40–85)	74.8 (11.4; 45.0–97.5)	76.7 (13.1; 42.5–97.5)

[a]Similar vowels exist in Mandarin.

group. For recent arrivals, only one significant correlation emerged: those who had spoken more English with their friends tended to perform better on the task, $r=0.31$, $p<0.01$. For past arrivals, better performance on the task was associated with a younger age at which English instruction began, $r=-0.55$, $p<0.001$, more years of U.S. education, $r=0.40$, $p<0.01$, and better English speaking ability of mothers, $r=0.42$, $p<0.05$. To further detect the unique predictive power of the four significant predictors for past arrivals, a hierarchical regression analysis was conducted. AoA and the age of English instruction were entered in the first step, followed by years of education in the U.S., and then the mother's English speaking ability. The two age variables accounted for 20% of the variance, $p<0.05$. Adding U.S. education did not change the amount of variance explained, but adding the mother's English speaking ability significantly increased it to 33%, $p<0.01$.

B. Production

The listeners showed high agreement rates on the produced vowel identity. Of the 1344 vowel tokens (168 participants $\times$ 8 vowels), five listeners agreed on 617 (45.90%) of the tokens. Another 331 (24.63%) tokens elicited agreement by four listeners. No judge showed obvious divergence from the group. The agreement rate varied among the vowels, ranging from 94% for /u/, to 41.67% for /ʌ/ by at least four listeners. This indicates that disagreements among the listeners were more likely due to the ambiguity of the productions rather than to listener factors. Taking these findings into account, data across all listeners were pooled together for analyses.

The production data from 42 participants in China, 50 recent arrivals, and 76 past arrivals were analyzed for both accuracy and error patterns. For accuracy analyses, all responses were scored as either correct or incorrect. When the intended vowel by the speaker and the chosen vowel by the listener matched, the response was correct. For each speaker, a percent correct score for a vowel was the proportion of correct responses out of five tokens. The total percent correct for all eight vowels was the average of the percent correct scores for the eight vowels.

1. Accuracy across groups and vowels

In this part of the analysis, performance accuracy, indicated by percent correct scores for all vowels and for each vowel were compared across the three participant groups. There was a wide range of accuracy levels across the different vowels (Table III). A mixed two-way 8 (vowels) $\times$ 3 (groups) ANOVA analysis revealed a main effect of group, $F(2,165)=14.36$ ($\eta^2=0.15$), a main effect of vowel, $F(6,1155)=37.76$ (ES=0.19; with the Greenhouse–Geisser correction of degrees of freedom), and an interaction between the group and vowel, $F(11,1155)=4.31$ ($\eta^2=0.05$; with the Greenhouse–Geisser correction of degrees of freedom) (all $p<0.001$).

The group effect reflects the finding that participants in China had a lower overall accuracy than both the recent and past arrivals. To further examine the group effect for each vowel, separate one-way ANOVA was performed for the individual vowel accuracy scores. There were significant group differences for /eɪ/ [$F(2,165)=28.82$, $p<0.001$], for /ɑ/ [$F(2,165)=9.0$, $p<0.001$], and for /ɪ/ [$F(2,165)=6.74$, $p<0.01$]. Bonferroni post-hoc tests revealed that, for all three vowels, participants in China scored significantly lower than both the recent and past arrivals, whereas these two latter groups were not significantly different from each other.

The main effect of vowel reflects varying performance on the vowels. The performance ranged from 94.99% for /u/ to 49.27% for /ʌ/. The interaction of group and vowel indicates that performance differences on vowels varied among the three groups. Visual inspection of Table III indicates that the two easiest vowels were /u/ and /i/ for all three groups, and the hardest two were /æ/ and /ʌ/. The recent and past arrivals had similar rank orders except a switch between ranks 2 and 3, and a switch between ranks 4 and 5. However, participants in China produced five vowels with accuracy around the lower 50% range, making the rank order less meaningful.

2. Age differences

Participants in China showed a nonsignificant correlation between age and average performance on all vowels, $r=0.26$, $p=0.10$. However, significant positive correlations between age and accuracy existed for two vowels, /eɪ/ [r

TABLE IV. Confusion matrix for the vowel productions by participants in China (first row; $n=42$), recent arrivals (second row; $n=76$), and past arrivals (third row; $n=50$).

Stimulus vowels (vowel said)	Response vowels (vowel heard)									
	i	ɪ	eɪ	ɛ	æ	ɑ	ʌ	u	ɔ	o/ʊ/ai
i	**88**	10	2.0							
	88	10	2.0							
	89	6	3					2		
ɪ	23	**52**	5	9	<1	<1	<1	7	<1	1
	16	**76**	4	3	<1		<1			<1
	12	**77**	4	6						1
eɪ	26	18	**54**							1
	2	1	**89**	1					<1	6
	<1	4	**88**	2						6
ɛ	1	8	<1	**65**	12	2	10			1
		3	2	**66**	20	2	5		<1	1
	<1	2	1	**72**	18	<1	5			2
æ				24	**52**	12	11			
		<1	2	23	**65**	7	1			
			1	22	**69**	5	2			1
ɑ					7	**51**	28		11	3
				<1	1	**77**	13		6	3
	<1			2	<1	**74**	17		6	
ʌ		<1		3	5	25	**54**		6	7
	<1			2	2	43	**45**		5	3
				<1	<1	40	**49**		6	4
u								**97**		3
						<1		**93**		7
	2							**95**		3

$=0.34$, $p<0.05$] and /ɛ/ [$r=0.34$, $p<0.05$]. No significant correlation was found for the recent arrivals. For the past arrivals, performance on two vowels, /ɪ/ [$r=-0.24$, $p<0.05$] and /eɪ/ [$r=-0.33$, $p<0.01$] showed significant negative correlation with AoA, a trend opposite that of the participants in China.

3. Error patterns

The overall error patterns were analyzed by creating confusion matrices for the three groups (Table IV). Responses were classified by the 8 target (intended) vowels contained in each of the /dVp/ utterances produced by participants, and by the 12 vowels given as the response alternatives. The numbers on a row indicate the percentage of instances an intended vowel (produced by all participants) was identified as one of the 12 vowels by the native listeners. The proportion of target and response matches (diagonal bold numbers on Table IV) was regarded as the accuracy score for each vowel.

The four vowels with the lowest accuracy rates (/ɛ, æ, ɑ, ʌ/) showed bidirectional confusion patterns, with the two vowels tested as discrimination pairs (/ɛ, æ/ and /ɑ, ʌ/) being highly confused with each other. However, although /ɛ/ or /æ/ were misidentified as each other in approximately equal proportions of the instances (17.4% and 22.8%, respectively), /ʌ/ was more often misheard as /ɑ/ (38%) than the opposite (18%). Vowels /u/ and /i/ had the highest accuracy scores. In between, /ɪ/ showed a concentrated confusion pattern, being most often heard as /i/. In contrast, /eɪ/ showed a more diffuse confusion pattern, heard as /i/, /ɪ/, or even /aɪ/. For both /ɪ/and /eɪ/, the immigrant groups showed considerable improvement in production accuracy.

TABLE V. Ranking of the bidirectional production error rate and discrimination accuracy for the six vowel pairs.

Vowel pairs	Production Bidirectional error rate/Difficulty rank		Perception Accuracy/Difficulty rank	
/i-ɪ/	8.5+16.3=24.8	3	94.93%	3
/i-eɪ/	2.2+7.8=10	5	98.20%	5
/ɛ-æ/	22.8+17.4=40.2	2	95.37%	4
/æ-ɑ/	7.9+2.5=10.4	4	87.62%	2
/ɑ-ʌ/	18+38=56	1	83.77%	1
/u-ɑ/	0.1+0=0.1	6	99.37%	6

C. Relation between perception and production

The relation between perception and production at both the individual level and group level was examined. The individual level relation was assessed by correlating perception and production total accuracy scores for the 168 native Mandarin speakers with measurable production data. There were significant positive correlations between perception and production performance for all participants together ($r=0.50$, $p<0.001$), for the participants in China ($r=0.42$, $p<0.001$), and for the past arrivals ($r=0.46$, $p<0.01$). The correlation for the recent arrivals was lower ($r=0.25$, $p=0.08$). The close to ceiling perception performance of the recent and past arrivals might have lowered the correlations.

At the group level, rank orders of difficulty in perception and production were compared (Table V). For production, we combined the performance for the eight vowels into the six vowel contrasts by obtaining the bidirectional confusion rates (summing up the percentage of time that one vowel was identified as the other for each pair) and ranking these numbers. For perception, we rank ordered the average perception correct scores for all the 168 participants who had production scores. The bivariate correlation of the two sets of ranked scores was marginally significant ($r=0.77$, $p=0.07$).

IV. DISCUSSION

In the current study we investigated how age-related differences in the perception and production of AE vowels changed with an increasing amount of AE exposure. We included three groups of native Mandarin speakers with varying amounts of L2 exposure: those with no L2 immersion experiences who represented a population of potential immigrants on their first day of arrival in the U.S., those with moderate L2 immersion experiences (in the U.S. for two years or less), and those with substantial L2 immersion experiences (in the U.S. for between three and five years). To assess the unique contribution of our focus variable, AoA, other potential predictor variables of L2 learning were also examined. The inclusion of AE vowels that bear different phonetic relations to Mandarin vowels permitted the investigation of the influence of L1 phonetic/phonological system on L2 phonological acquisition. The findings add to a more

accurate description of age-related differences in L2 phonological learning, and call for a more refined theoretical account of the phenomenon.

With increasing L2 use, age differences in performance accuracy changed from an older-learner advantage to a younger-learner advantage for both perception and production. For the participants in China with no L2 immersion experiences, an older chronological age predicted a significantly higher discrimination accuracy of all vowel contrasts and higher production accuracy of two difficult vowels.[2] For the recent arrivals, AoA was not related to performance at all. For the past arrivals, a younger AoA predicted significantly better discrimination accuracy for three vowel contrasts, and better production accuracy for two vowels.

The interaction of age-related differences with the amount of L2 exposure is consistent with the earlier study that demonstrated this full crossover pattern (Snow and Hoefnagel-Höhle, 1977). Notably, the findings of the current study and that of Snow and Hoefnagel-Höhle were obtained from different language populations (Mandarin-English versus English-Dutch), with different time sampling methods (cross-sectional versus longitudinal), and different linguistic foci (vowel perception and production in nonsense disyllables versus real word perception and production). These findings further strengthen the view that older learners (or later arrivals in the immigration setting) initially have an advantage over younger learners (or early arrivals in the immigration setting), but this advantage disappears and then becomes reversed over the course of L2 immersion.

In light of these findings, theories that address age-related differences in phonological learning must explain not only the long-term younger learner advantage (as is the traditional focus), but also the short-term older-learner advantage and the processes of change involved. Although all three theoretical accounts predict and explain the long-term younger learner advantage, they are not similarly powerful in explaining the age-related differences exhibited prior to a long-term time point.

The Critical/Sensitive period hypothesis faces a challenge to explain why the genetically preprogrammed advantage of younger learners takes time to exert its effect. In light of the current findings, the theory should at least specify that, whatever the advantage younger learners have in phonological learning, it does not include an immediate superiority in perceptual sensitivity to and production accuracy of new speech sounds. In addition, this theory needs to specify domain specific phonological learning mechanisms that favor children within a sensitive period, and that take effects cumulatively over time.

The Environmental theory is fully consistent with our findings of the changing age differences with increasing L2 immersion. According to this theory, early and late arrivals in the immigration setting experience different language environments. In a longitudinal study of native Mandarin speakers learning English in the U.S., Jia and Aaronson (2003) found that early arrivals were exposed to a significantly richer L2 environment than late arrivals. The richer L2 environment was not only indicated by a greater quantity of L2 exposure, but also by a higher quality of L2 exposure, such as the intensive and highly motivated use of L2 during interactions with peers and the dominant society culture. Such language environment differences can accumulate over the years of L2 immersion, and lead to proficiency advantages of early over late arrivals.

Three types of findings from the current and other studies support the Environmental account. First, consistent with findings from the longitudinal study (Jia and Aaronson, 2003), the current results indicated that a younger AoA was associated with significantly more L2 use with parents, siblings, and friends averaged over the years of L2 immersion. Second, if the younger-learner advantage arises from a richer L2 environment, an early age of exposure should not guarantee high proficiency if the L2 environment is not rich enough. Flege *et al.* (1999) found that among native Italian speakers who immigrated to Canada at a young age, those who used English more were significantly better in discriminating English vowels than those who used English less. Third, if a richer L2 environment benefits early learners, more L2 exposure should benefit all learners. In the current study, native Mandarin speakers of all ages made significant improvement in AE vowel perception and production through L2 immersion. Further, recent arrivals who used more English with their friends, and past arrivals whose mothers had higher English-speaking abilities, discriminated the AE vowel pairs more accurately than the others.

The L1 Transfer/Interference theory postulates that older L2 learners' difficulty is due to their more developed abilities in L1. Although the L1 phonological system is acquired early in life, speech motor control (e.g., Kent, 1992) and perceptual representations (e.g., Walley and Flege, 1999) for L1 speech sounds develop gradually through childhood and into adolescence. Along with these developments, older children and adults become more likely to perceive and produce new speech sounds through the filter of their L1 phonological system. In comparison to their older counterparts, native Korean-speaking children were less likely to assign Korean vowel categories to English vowels, suggesting that "children's L1 sound categories are less powerful attractors of L2 sounds than those of adults" (Baker *et al.*, 2002, p. 41). Such an account should predict a younger-learner advantage from the beginning of L2 immersion. Therefore, findings of the initial older-learner advantage of the current study and others (e.g., Politzer and Weiss, 1969; Snow and Hoefnagel-Höhle, 1977) do not appear to be in line with this account. However, a possible explanation exists. In the beginning, the L1 interference/transfer effects among older learners are not significant enough to override the overall perceptual and performance advantage they have in an optimal processing situation. The phonetic features detected by older learners may not be utilized during on-line processing of speech (Flege, 2003), and thus not contribute to the growth of proficiency.

If a more established L1 system renders late learners at a disadvantage for L2 speech learning, we should also expect that the acquisition of L1 has an impact on any speakers of L1, regardless of the age. Therefore, our findings of specific

influences of the L1 vowel system on L2 vowel learning serves as indirect evidence for the L1 Transfer/Interference account. Difficulty rankings for perception of vowel contrasts and production of vowels were similar across the three participant groups. For perception, the order of difficulty closely reflected the hypothesized order based on both phonetic similarity and hypothesized perceptual assimilation patterns influenced by L1 vowel space (Best, 1995). In the two most difficult pairs /ɛ-æ/ and /ɑ-ʌ/, the two vowels involved in each do not have close counterparts in Mandarin (not including allophonic variations), and are close in vowel space. Larger acoustic distances (i.e., /æ-ɑ/) or the presence of one of the two vowels in Mandarin (i.e., /i-e^{I}/, /i-I/) was associated with medium level performance. Similarly for vowel production, /ɛ, æ/, with no close Mandarin counterparts, showed symmetrical confusions. AE /ɑ, ʌ/ were also confused in production, although the confusions were asymmetrical favoring /ɑ/. The two vowels that had corresponding LI counterparts /i, u/ showed close to ceiling accuracy in intelligibility, even when produced by speakers with no L2 immersion experiences.

The current study yielded a positive correlation between perception and production at the individual and group levels. At the individual level, better perception performance significantly predicted better production performance. At the group level, the vowel contrasts that were harder to distinguish in the perception task also had the highest bidirectional confusion rate in production. Vowel contrasts that were better distinguished were also produced with greater accuracy. These findings are consistent with those in the literature. Flege and colleagues found similar positive correlations between English vowel intelligibility and discrimination among native speakers of various languages (Flege *et al.*, 1997; 1999). For example, both native Korean and Mandarin speakers identified synthetic vowels along the bat-bet (/æ-ɛ/) and beat-bit (/i-ɪ/) continua differently from native English speakers, and produced the two vowels in a contrast with bidirectional confusion (Flege *et al.*, 1997). We note that the nature of such a positive relation between perception and production is still controversial. According to SLM, accurate L2 production to a large extent relies on accurate perception, and thus, perception development should precede production (Flege, 1995; McAllister, Flege, and Piske, 2002). Other researchers emphasize the causal role of production in perception. For example, Japanese speakers' production of English /r/ and /l/ was more accurate than their perception (e.g., Sheldon and Strange, 1982). However, perceptual training on the /r-l/ contrast did lead to production improvement by Japanese speakers (Bradlow, Akahane-Yamada, Pisoni, and Tohkura, 1999). Notably, tasks measuring production and perception abilities may be inherently incommensurable (e.g., Flege, 1999; Tsukada *et al.*, 2005). They can pose varying levels of processing demands by the choice of stimuli, tasks, and procedures. The current production task promoted optimal performance with minimal processing demands. Only one consonantal context was used, and productions were rated in terms of intelligibility rather than degree of foreign accent (the latter is more stringent than the former, as found in studies such as Munro *et al.*, 1996). Given this, several of our findings are consistent with the predictions of the SLM that production abilities at least partially rely on perception abilities. First, some Mandarin speakers were able to distinguish some vowel pairs accurately but confused them in production, indicating that production indeed lagged behind perception. Second, perception abilities improved at a faster rate than production abilities.

There are several limitations of the current study that can be addressed by future research. First, we manipulated the length of L2 exposure cross-sectionally rather than longitudinally. Some aspects of the participants in China and those in the U.S. were not completely comparable, such as the testing environments and Chinese dialect backgrounds. Nevertheless, the incomparability between the groups would have mainly affected our interpretations of the between group differences, not the age trends within each group. Second, in order to access optimal performance, the current study minimized speaker and contextual variations of the stimuli as well as the processing demands of the tasks. Future research should increase the variations along these dimensions to more closely approximate "on-line" phonological processing and learning.

In summary, the current findings indicate that age and amount of L2 immersion jointly influenced learning, indicated by a dynamic change of age-related differences with increasing exposure to L2. These findings support a combined Environmental and L1 Interference/Transfer theory as an explanation for the long-term younger-learner advantage in mastering L2 phonology. With increasing age, the growing influence of L1 perception and production patterns, coupled with L2 input of lesser quantity and quality, leaves the long-term achievement in L2 phonology of most older arrivals behind that of the younger arrivals. Our findings also indicate that older learners have their unique advantages in nonnative speech learning. Future research should investigate how the strengths and weaknesses of younger and older learners interact in the learning processes, and tailor L2 speech learning and training strategies to learners of all ages.

ACKNOWLEDGMENTS

Research reported in this article was supported by a City University of New York collaborative grant (No. 80209-01-08) awarded to the first and second authors, and a grant from the National Institute of Health (SCORE/NICHD #41353-11-19/20) to the first author. We thank Alan Wu for collecting the data from participants in New York City. We are grateful to the following colleagues who offered methodological advice at various points of the project, or read drafts of the manuscript: Jim Jenkins, Richard Bock, Akiko Fuse, Anne Reid, Keith Happaney, Erika Levy, Kanea Nishi, Puisan Wang and Shari Berkowitz. We also thank Bruno Tagliaferri for his technical support.

APPENDIX: ACOUSTICAL CHARACTERISTICS OF THE VOWEL STIMULI (AVERAGE VALUES OF THE THREE TOKENS FOR EACH VOWEL)

Stimuli	Duration	F1 (Hz)	F2 (Hz)
/i/	122	279.6	2679.4
/ɪ/	93.4	482.5	2192.6
/e[1]/	157.3	529.1	2438.1
/ɛ/	127.3	723.8	2107.9
/æ/	176.2	969.3	1866.7
/ɑ/	168.0	1007.4	1451.8
/ʌ/	128.2	876.2	1540.7
/u/	121.0	304.8	1557.7

Note. Acoustic values for Mandarin vowels in dVp context are not available, and those for Mandarin vowels in contexts not identical to that of AE are not included in this table because direct comparison between vowels in different contexts are inappropriate (Strange *et al.*, 1976).

[1]In this article, immersion refers to the holistic setting of learning a second language, L2, in the immigration setting, both in and outside school. It carries a different meaning from the concept of immersion in education, which refers to L2 acquisition through natural contact with L2 outside school as opposed to classroom instruction.

[2]The older learner advantage in perception is likely due to multiple factors. First, it is partially due to growth in language related speech perception abilities. Studies of auditory perception development have shown a significant growth in the perceptual sensitivity to elements of speech signals related to language, such as signal duration (e.g., Elfenbein, Small, and Davis, 1993; Jensen and Neff, 1993). Second, it is partially due to growth of perceptual sensitivity that is not specific to cross-language phonetic influence. Tsukada *et al.* (2005) found that monolingual English-speaking children scored significant lower than monolingual-English speaking adults in discriminating /ɑ-ʌ/, but not /æ-ɛ/, /i-ɪ/, and /e[1]-ɛ/. Third, it is partially due to more years of English education, as partialling out years of English education reduced the significant growth trend to a marginally significant level. The older learner advantage is unlikely to be due to disadvantages of younger children to handle the tasks. Younger children passed the practice Mandarin trials in a similar proportion as older participants did. The task duration was only 10–15 min. Younger children performed equally well on the six blocks, showing no indication of attention decline.

Aoyama, K., Flege, J. E., Cuion, S. G., Akahane-Yamade, R., and Yamada, T. (**2004**). "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/," J. Phonetics **32**, 233–250.

Asher, J., and Garcia, R. (**1969**). "The optimal age to learn a foreign language," Mod. Lang. J. **53**, 334–341.

Baker, W., Trofimovich, P., Mack, M., and Flege, J. (**2002**). "The effect of perceived phonetic similarity on non-native sound learning by children and adults," in *Proceedings of the 26th Annual Boston University Conference on Language Development*, edited by B. Skarabela, S. Fish, and A. Do (Cascadilla P., Somerville, MA), pp. 36–47.

Best, C. (**1995**). "A direct realistic perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York P, Baltimore, MD), pp. 171–206.

Beving, B., and Eblen, R. E. (**1973**). '"Same" and "different" concepts and children's performance on speech sound discrimination,' J. Speech Hear. Res. **16**, 513–517.

Bohn, O. S., and Flege, J. E. (**1992**). "The production of new and similar vowels by adult German learners of English," Stud. Second Lang. Acquis. **14**, 131–158.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," Percept. Psychophys. **61**, 977–985.

Cochrane, R. (**1980**). "The acquisition of /r/ and /l/ by Japanese children and adults learning English as a second language," J. Multiling. and Multicul. Dev. **1**, 331–360.

Cochrane, R., and Sachs, J. (**1979**). "Phonological learning by children and adults in a laboratory setting," Lang Speech **22**, 145–149.

Elfenbein, J. L., Small, A. M., and Davis, J. M. (**1993**). "Developmental patterns of duration Discrimination," J. Speech Hear. Res. **36**, 842–849.

Fathman, A. (**1975**). "The relationship between age and second language learning productive ability," Lang. Learn. **25**, 245–253.

Flege, J. E. (**1995**). "Second language speech learning theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York P., Baltimore, MD), pp. 233–277.

Flege, J. E. (**1999**). "The relation between L2 production and perception," in *Proceedings of the 14th International Congress of Phonetic Sciences*, edited by J. Ohala, Y. Hasegawa, M. Ohala, D. Granveille, and A. Bailey (Department of Linguistics, University of California at Berkeley, Berkeley, CA), pp. 1273–1276.

Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin, Germany), pp. 319–355.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., and Tsukada, K. "Degree of foreign accent in English sentences produced by Korean children and adults," J. Phonetics (in press).

Flege, J. E., Bohn, O. S., and Jang, S. (**1997**). "Effects of experience on non-native speakers' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J. E., MacKay, I., and Meador, D. (**1999**). "Native Italian speakers' production and perception of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995a**). "Effects of age of second-language learning on the production of English consonants," Speech Commun. **16**, 1–26.

Flege, J. E., Munro, M. J., and MacKay, I. R. A. (**1995b**). "Factors affecting strength of perceived foreign accent in a second language," J. Acoust. Soc. Am. **97**, 3125–3134.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111

Howie, J. M. (**1976**). "*Acoustical Studies of Mandarin Vowels and Tones*" (Cambridge University Press, New York).

Jensen, J. K., and Neff, D. L. (**1993**). "Development of basic auditory discrimination in preschool children," Psychol. Sci. **4**, 104–107.

Jia, G., Aaronson, D., and Wu, Y. H. (**2002**). "Long-term language attainment of bilingual immigrants: Predictive factors and language group differences," Appl. Psycholing. **23**, 599–621.

Jia, G., and Aaronson, D. (**2003**). "A longitudinal study of Chinese children and adolescents learning English in the US," Appl. Psycholing. **24**, 131–161.

Kent, R. D. (**1992**). "The biology of phonological development," in *Phonological Development: Models, Research, Implications*, edited by C. A. Ferguson, L. Menn, and C. Stoel-Gammon (York P., Timonium, MD), pp. 65–90.

Kewley-Port, D., Akahane-Yamada, R., and Aikawa, K. (**1996**). "Intelligibility and acoustical correlates of Japanese accented English vowels," *Proceedings of Int. Conference on Spoken Lang. Process*. Vol. **96**, pp. 450–453.

Kuhl, P. (**2000**). "A new view of language acquisition," Proc. Natl. Acad. Sci. U.S.A., **97**, 11850–11857.

Lin, Y. H. (1989). "Autosegmental treatment of segmental process in Chinese phonology," unpublished Ph.D. thesis, University of Texas at Austin.

MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholing. **2**, 369–390.

MacKay, I. R. A., Flege, J. E., Piske, T., and Schirru, C. (**2001**). "Category restructuring during second-language speech acquisition," J. Acoust. Soc. Am. **110**, 516–528.

McAllister, R., Flege, J. E., and Piske, T. (**2002**). "The influence of L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian," J. Phonetics **30**, 229–258.

Munro, M., Flege, J., and MacKay, I. (**1996**). "The effect of age of second language learning on the production of English vowels," Appl. Psycholing. **17**, 313–334.

Oyama, S. (**1976**). "A sensitive period for the acquisition of a nonnative phonological system," J. Psycholing. Res. **5**, 261–283.
Patkowski, M. S. (**1990**). "Age and accent in a second language: A reply to James Emil Flege," Appl. Linguist. **11**, 73–89.
Peterson, G. E., and Barney, H. L. (**1952**). "Control method used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Politzer, R. L., and Weiss, L. (**1969**). "Developmental aspects of auditory discrimination, echo response and recall," Mod. Lang. J. **53**, 75–85.
Rogers, C. L. (1997). "Intelligibility of Chinese-accented English," unpublished Ph.D. thesis, Indiana University, Bloomington.
Scovel, T. (2000). "A critical review of the critical period research," Ann. Review Appl. Ling. **20**, 213–223.
Sheldon, A., and Strange, W. (**1982**). "The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception," Appl. Psycholing. **3**, 243–261.
Snow, C. E. (**1983**). "Age differences in second language acquisition: Research findings and folk psychology," in *Second Language Acquisition Studies*, edited by K. M. Bailey, M. H. Long and S. Peck (Newbury House, Rowley, MA), pp. 141–150.
Snow, C. E., Hoefnagel-Höhle, M. (**1977**). "Age differences in the pronunciation of foreign sounds," Lang. Speech **20**, 357–365.
Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K., and Jenkins, J. (**1998**). "Perceptual assimilation of American English vowels by Japanese listeners," J. Phonetics **26**, 311–344.
Strange, W., Verbrugge, R. R., Shankweiler, D. P., and Edman, T. R. (**1976**). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. **79**, 1086–1100.
Tahta, S., Wood, M., and Loewenthal, K. (**1981a**). "Age changes in the ability to replicate foreign pronunciation and intonation," Lang Speech **24**, 363–372.
Tahta, S., Wood, M., and Loewenthal, K. (**1981b**). "Foreign accents: Factors relating to transfer of accent from the first language to a second language," Lang. Speech **42**, 265–272.
Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., and Flege, J. E. (**2005**). "A developmental study of English vowel production and perception by native Korean adults and children," J. Phonetics **33**, 263–290.
Walley, A. C., and Flege, J. E. (**1999**). "Effect of lexical status on children's and adults' perception of native and non-native vowels," J. Phonetics **27**, 307–332.
Wan, I. P. (1999). "Mandarin phonology: Evidence from speech erros," unpublished Ph.D. thesis, State University of New York at Buffalo.
Werker, J. F., and Tees, R. C. (**1999**). "Influences on infant speech processing: Toward a new synthesis," Annu. Rev. Psychol. **50**, 509–535.
Yeni-Komshian, G., Flege, J. E., and Liu, S. (**2000**). "Pronunciation proficiency in the first and second languages of Korean-English bilinguals," Biling: Lang. Cog. **3**, 131–149.

Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners[a)]

Mini N. Shrivastav,[b)] Larry E. Humes, and Diane Kewley-Port
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 26 May 2005; revised 28 October 2005; accepted 21 November 2005)

Speech-understanding difficulties observed in elderly hearing-impaired listeners are predominantly errors in the recognition of consonants, particularly within consonants that share the same manner of articulation. Spectral shape is an important acoustic cue that serves to distinguish such consonants. The present study examined whether individual differences in speech understanding among elderly hearing-impaired listeners could be explained by individual differences in spectral-shape discrimination ability. This study included a group of 20 elderly hearing-impaired listeners, as well as a group of young normal-hearing adults for comparison purposes. All subjects were tested on speech-identification tasks, with natural and computer-synthesized speech stimuli, and on a series of spectral-shape discrimination tasks. As expected, the young normal-hearing adults performed better than the elderly listeners on many of the identification tasks and on all but two discrimination tasks. Regression analyses of the data from the elderly listeners revealed moderate predictive relationships between some of the spectral-shape discrimination thresholds and speech-identification performance. The results indicated that when all stimuli were at least minimally audible, some of the individual differences in the identification of natural and synthetic speech tokens by elderly hearing-impaired listeners were associated with corresponding differences in their spectral-shape discrimination abilities for similar sounds. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151794]

PACS number(s): 43.71.Ky, 43.66.Sr [GDK] Pages: 1131–1142

I. INTRODUCTION

Elderly hearing-impaired listeners often experience speech-understanding difficulties, frequently greater in magnitude than predicted by the degree of hearing loss (CHABA, 1988; Marshall, 1981). It is generally agreed that the primary factor contributing to speech-understanding difficulties in the elderly hearing-impaired is the audibility of the speech signal, which in turn is related to the degree and configuration of hearing loss (Humes and Roberts, 1990; Humes, Watson, Christensen, Cokely, Halling, and Lee, 1994; van Rooij and Plomp, 1990a, 1990b, 1992; van Rooij *et al.*, 1989). Other factors, such as spectral and temporal processing deficits, cognitive deficits, and central auditory deficits, have all been found to make more limited contributions to speech-understanding difficulties in these listeners (CHABA, 1988; Gordon-Salant and Fitzgibbons, 1997; Jerger, Jerger, Oliver, and Pirozzolo, 1989; van Rooij and Plomp, 1990b, 1992; van Rooij *et al.*, 1989).

To address these difficulties in unaided speech understanding, the most widely used rehabilitation strategy for elderly hearing-impaired persons is amplification. Approximately two-thirds of the hearing-aids sold in 2000, for example, were purchased by individuals over 65 years of age (Skafte, 2000). However, even with well-fit hearing aids, many of these individuals still seem to experience considerable speech-understanding difficulties (Plomp, 1978). In addition, there is considerable individual variability in speech-understanding performance, even among elderly hearing-impaired individuals who have nearly identical audiometric profiles and are fitted with similar amplification devices. The reasons for this variability are not clear.

Various studies have found that suprathreshold auditory processing factors have little, if any, effect on recognition of amplified speech in elderly hearing-impaired listeners (e.g., Humes *et al.*, 1994; Humes, 2002). Individual differences in cognitive abilities might account for a part of the variability in aided speech-recognition performance across listeners (Humes, 2002). However, there is evidence that age-related factors other than audibility might contribute to speech-understanding difficulties in tasks complicated by factors such as multiple degradations in stimuli, multiple talkers, low contextual cues, and greater memory load (Gordon-Salant and Fitzgibbons, 1995a, 1995b, 1997).

Speech-understanding difficulties observed in elderly hearing-impaired listeners seem to be associated with problems in the identification of consonants rather than vowels (e.g., Flynn, Dowell, and Clark, 1998). These listeners have the greatest difficulty identifying consonants from the stop and fricative manner-of-articulation categories (Gordon-Salant, 1987). Within these two manner categories, errors, most commonly in the place of articulation, result in significant speech-understanding difficulties (Flynn *et al.*, 1998;

[a)]Portions of this work were presented at the 148th meeting of the Acoustical Society of America, San Diego, California, November 2004.

[b)]Currently at: Department of Communication Sciences and Disorders, University of Florida, Gainesville, Florida 32611. Electronic mail: mnarendr@csd.ufl.edu

Gelfand, Piper, and Silman, 1986). For example, the voiceless stop consonants /p/, /t/, and /k/, and the voiceless fricatives /f/, /s/, and /ʃ/, differing only in place of articulation, are highly confusable for elderly hearing-impaired listeners when compared to elderly normal-hearing listeners. (Gordon-Salant, 1987). In the case of stop consonants, place of articulation is cued primarily by the shape of the burst spectra and formant transitions (Stevens and Blumstein, 1978; Kewley-Port, Pisoni, and Studdert-Kennedy, 1983). For fricatives, place of articulation is cued primarily by the frequency range and shape of the frication-noise spectrum (Heinz and Stevens, 1961). Thus, spectral shape is an important feature that distinguishes place of articulation within a particular manner-of-articulation class, such as stops or fricatives.

The contribution of spectral shape as an acoustic cue is not limited to speech; rather, it is a vital cue for nonspeech sounds like music and commonly occurring environmental sounds (Moore, 2003). Consequently, it would seem that the ability to identify and discriminate spectral shape would be important for recognition of speech and other environmental sounds as well. Most previous studies of spectral-shape discrimination primarily have involved multitonal complexes used in profile-analysis experiments (Green, 1988). These studies use some degree of amplitude roving to ensure that listeners are basing their responses on spectral shape and not overall level differences within a trial. However, these stimuli do not resemble many consonant sounds of speech. Further, only a few studies have included hearing-impaired listeners. This is due, in part, to methodological problems such as the difficulty in ensuring audibility of all components of the spectrum and using amplitude rove in light of the reduced dynamic range of hearing-impaired listeners. Summers and Leek (1994) measured higher thresholds for hearing-impaired listeners when compared to normal-hearing listeners on a discrimination task involving 160-tone equal-amplitude and rippled spectra, but concluded that the poor spectral-shape discrimination ability of hearing-impaired listeners was related to the widened auditory filters in these listeners. Lentz and Leek (2002) measured thresholds for a signal added to the central component of a 300-ms five-tone complex presented with and without perturbation and amplitude rove for a group of six normal-hearing (age 28–68 years) and six elderly hearing-impaired listeners (age 65–76 years). The stimuli were presented at levels high enough to ensure audibility for all listeners. The results indicated that for both groups of listeners, thresholds were higher for the rove and perturbation conditions. Based on a cue-weighting analysis, the authors concluded that listeners might have based their responses on the change in level of the signal component rather than using simultaneous comparisons across multiple channels. Lentz and Leek (2002; 2003) found a great deal of individual variability across all listeners. Also, they concluded that a few of the hearing-impaired listeners were using weighting strategies different from the normal-hearing listeners.

The present study was aimed at investigating whether the ability to identify and discriminate spectral shape is related to individual differences in speech-identification performance among elderly individuals when audibility of the full bandwidth of the speech signal has been ensured. Spectral shape discrimination was measured for speechlike and nonspeech stimuli with the use of amplitude rove following equalization of rms amplitudes. Speech-identification was measured for both natural (consonant-vowel or vowel-consonant nonsense syllables) and synthesized (stop bursts and fricative noise spectra) tokens similar to consonant sounds in speech. A group of young normal-hearing adults was also included to provide reference measures for performance on these tasks.

II. METHODS

A. Participants

The study involved two groups of listeners: (1) a group of five male and five female young normal-hearing adults (YNH), ranging in age from 23 to 35 years (M = 27.9 years); and (2) a group of ten male and ten female elderly hearing-impaired (EHI) adults ranging in age from 66 to 85 years (M = 74.7 years). None of the YNH subjects had a history of hearing loss and all had air-conduction thresholds less than or equal to 20 dB HL (ANSI, 1996) from 250 through 4000 Hz. All the EHI listeners had hearing loss of sensorineural origin and none of them had a history of fluctuating hearing loss or middle-ear pathology. The means and range of air-conduction hearing thresholds for the test ear of the 20 EHI subjects are shown as the filled triangles and associated error bars in Fig. 1.

B. Stimuli

The stimuli used for speech identification will be described first. This will be followed by descriptions of the stimuli for spectral-shape discrimination.

1. Speech identification

Speech identification was tested using two groups of stimuli: (1) The City University of New York (CUNY) Nonsense Syllable Test (NST) (Levitt and Resnick, 1978); and (2) a set of computer-synthesized fricatives and stops. Each of these stimulus sets is described, in turn, below.

a. NST: The CUNY NST is a standardized speech-identification test involving the closed-set identification of a group of nonsense syllables. The original NST includes 91 items organized into 11 subtests, each comprised of 7–9 nonsense syllables. Within a given subtest of the NST, a group of voiced ("C") or voiceless ("c") consonants is tested in the same vowel context (either /i/, /u/ or /a/) and syllable format (consonant-vowel, CV, or vowel-consonant, VC). The consonants tested include voiced and voiceless stops, fricatives, affricates, nasals, liquids, and glides. Based on this categorization, the 11 subtests, in the order of testing in the present study, were labeled "Ca1," "cu," "Ca2," "ic," "iC," "ca," "ac," "aC," "uC," "uc," and "ci." Subtests 1 and 3 were labeled Ca1 and Ca2 as both involved initial voiced consonants followed by the vowel /a/. The specific syllables comprising each of these 11 subtests are presented in Table I. The original NST presents these syllables in a carrier phrase "You will mark __ please." The syllables used in this experiment

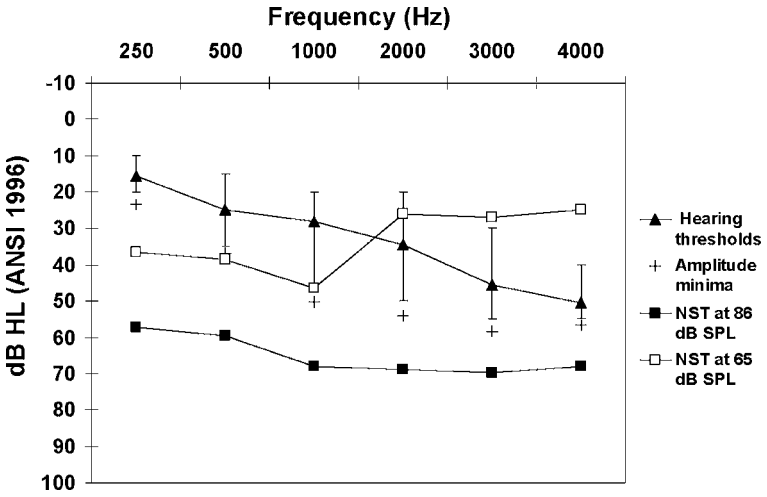


FIG. 1. Mean and range of thresholds for the test ears of the 20 EHI subjects are shown using closed triangles and vertical lines, respectively. The open squares show one-third octave band levels for the original unshaped NST stimuli for an overall presentation level of 65 dB SPL. The filled squares show the one-third octave band levels for the shaped NST stimuli presented at an overall level of 86 dB SPL (see "Stimuli" for details). A calibration noise that represented the amplitude minima for all stimuli across frequency was used to establish the maximum permissible amounts of hearing loss. These amplitude minima are shown using "+" symbols. These values were used to define the maximum amount of permissible hearing loss at each frequency for the hearing-impaired subjects, such that the stimuli representing the minimum amplitude would be above threshold at each frequency.

were excised from the carrier phrase and digitized with 16-bit resolution and a 48 828-Hz sampling rate.

Overall and one-third octave-band levels were measured for these stimuli using ER-3A insert earphones in a 2-cm^3 coupler. Based on these measures, one-third octave-band levels for a conversational level (65 dB SPL) were obtained, and these are shown as the open squares in Fig. 1. It is apparent that at this level, much of the high-frequency energy in these stimuli would be inaudible to the hearing-impaired listeners. In order to ensure audibility of the stimuli across the spectrum up to 5 kHz without overamplifying the low frequencies, the NST spectrum was shaped such that a gain of 43 dB and 21 dB was provided for frequencies between 1600–5000 Hz and 200–1000 Hz, respectively. Sound intensity below 200 Hz and above 5000 Hz was attenuated using a bandpass filter. One-third octave band levels for the resulting shaped NST stimuli are shown in Fig. 1 using the filled squares. These shaped NST stimuli were used for the identification testing, and were presented to the listeners at an overall level of 86 dB SPL.

b. Synthesized fricatives and stops: The spectra of these stimuli resembled the spectra of the frication noise of voiceless fricatives /f/, /s/, and /ʃ/, and the spectra of the onset of voiceless stops /p/, /t/, and /k/. These stimuli were created using the Johnson (1987) version of the Klatt speech synthesizer (Klatt, 1980). All the voiceless fricative and stop tokens in this study were created by modifying the frequency, amplitude and bandwidth of six formants represented by resonators of the synthesizer in the parallel configuration.

The frication spectrum of /f/ was represented in the synthesizer by providing 55 dB gain for the bypass path (AB) and 0 dB gain on each of the six formants. The overall gain (Go) was 65 dB, and the amplitude of frication (AF) was 60 dB. The frication noise of /s/ was synthesized by modifying the amplitude spectrum of /f/ to reflect a gain of 60 dB at 4900 Hz, the sixth-formant center frequency. Similarly, the frication noise of /ʃ/ was synthesized by modifying the spectrum of /f/ to reflect a gain of 60 dB at 2300 Hz, the third-formant center frequency. These stimuli were resampled to 48 828 Hz from the original 11 025 Hz, and spectra representing 2048-point FFTs with Blackmann windowing were computed from the digitized stimulus files. The /s/ stimulus was much more intense than /ʃ/, and /f/ was the weakest. In order to minimize the contribution of overall amplitude cues for the identification of these sounds, stimulus waveforms were adjusted in overall amplitude to yield equivalent rms amplitudes. In addition, all the stimuli were low-pass filtered with a cut-off frequency of 5000 Hz. This was done to maintain uniformity across the different types of stimuli used for identification and discrimination. Further, it was necessary to restrict the frequency range of the stimuli because of the sloping high-frequency hearing loss of the elderly listeners. The resulting amplitude spectra are shown in the top, middle, and bottom panels, respectively, of Fig. 2. (Note that the sound represented by the phonetic symbol /ʃ/ will be referred to as /sh/ in figures.) It is important to note that other than the amplitudes of the sixth and third formant, respectively, no other synthesis parameters of /f/ were

TABLE I. The individual subtests of the NST in IPA format.

Subtest	Test syllables
Ca1	/ba/, /va/, /za/, /da/, /θa/, /na/, /ga/, /ma/
cu	/tʃu/, /hu/, /ʃu/, /θu/, /ku/, /fu/, /pu/, /tu/, /su/
Ca2	/da/, /wa/, /ba/, /ra/, /ja/, /la/, /ga/, /dʒa/
ic	/ip/, /is/, /ik/, /iθ/, /it/, /if/, /iʃ/
iC	/iθ/, /ig/, /im/, /id/, /iv/, /iŋ/, /ib/, /iz/, /in/
ca	/fa/, /sa/, /pa/, /θa/, /ka/, /tʃa/, /ha/, /ʃa/, /ta/
ac	/ɔf/, /ɔp/, /ɔʃ/, /ɔt/, /ɔθ/, /ɔk/, /ɔs/
aC	/ɔn/, /ɔd/, /ɔz/, /ɔv/, /ɔg/, /ɔm/, /ɔθ/, /ɔb/, /ɔŋ/
uC	/un/, /uz/, /ub/, /uŋ/, /uv/, /ud/, /um/, /ug/, /uθ/
uc	/uʃ/, /uk/, /us/, /up/, /uf/, /ut/, /uθ/
ci	/ti/, /ʃi/, /pi/, /θi/, /fi/, /si/, /ki/, /hi/, /tʃi/

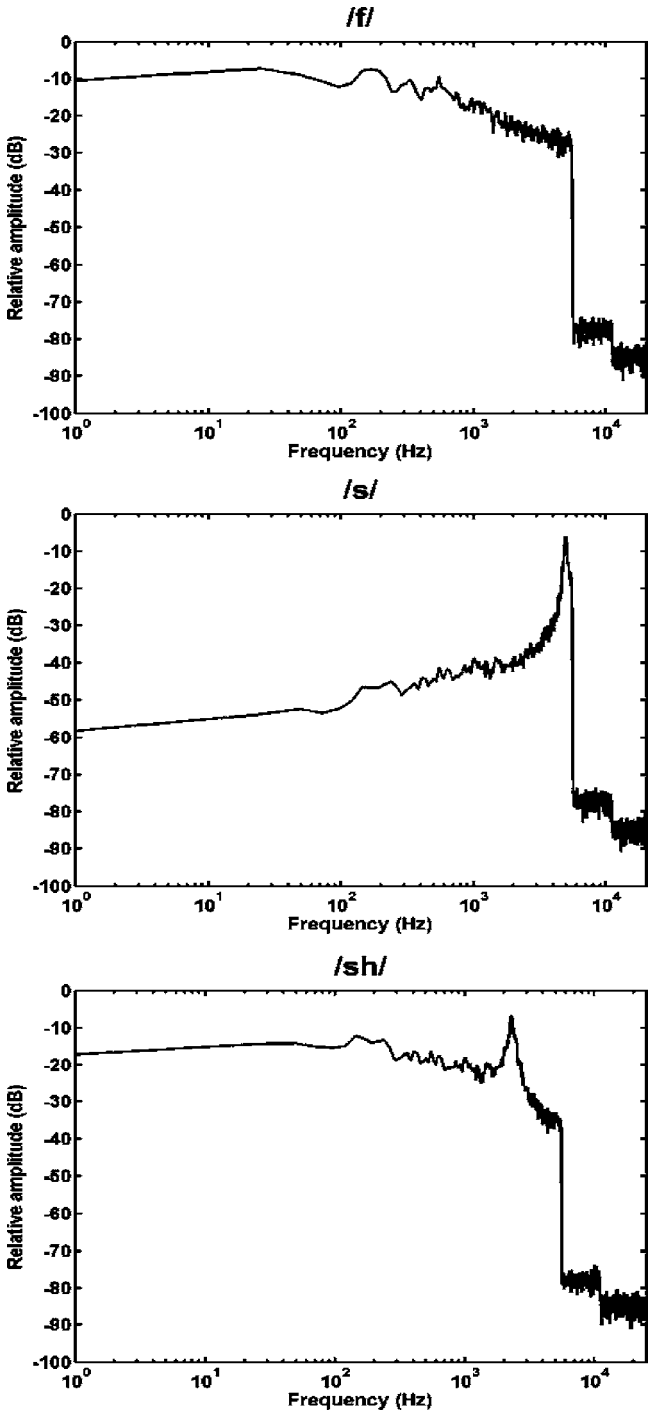


FIG. 2. Amplitude spectra of the three 250-ms fricative stimuli. The symbol /sh/ in the figure stands for the sound represented by the phonetic symbol /ʃ/. These spectra were computed from the digitized, resampled (sampling rate of 48 828 Hz), and low-pass filtered stimulus files and represent 2048-point FFTs with Blackmann windowing. Stimulus waveforms were adjusted to yield equivalent rms amplitudes.

changed during the synthesis of /s/ and /ʃ/. Stimulus duration for each of these three stimulus tokens was 250 ms, including a 10-ms onset and offset ramp.

The onset spectra of /p/, /t/, and /k/ are similar to the frication noise spectra of the fricatives /f/, /s/, and /ʃ/, respectively, but the stimuli are much shorter in duration. Hence, the stop-consonant stimuli used in the study were created by cutting 10-ms segments at the zero crossings from the corresponding 250-ms fricative spectra. It was necessary to use stimuli as short as 10 ms because preliminary informal listening experiments revealed that the stimuli began to sound like affricates, rather than stop bursts, at longer durations. Due to their short durations, the onsets and offsets of these stimuli were not ramped. Inspection of the amplitude spectra for these shortened stimuli showed some spectral splatter due to the short duration, but the general shape of spectra illustrated previously in Fig. 2 was retained. This was confirmed by spectral analysis of a concatenated series of 20 of the 10-ms stimuli.

2. Discrimination testing

Discrimination testing was performed for two different types of stimuli: (1) the synthesized fricative and stop tokens described earlier for the identification task; and (2) a set of spectrally tilted broadband noise stimuli. These two sets of stimuli are described below.

a. Synthesized fricative and stop stimuli: The stimuli in this task were based on the "prototypical" fricative and stop tokens used in the fricative- and stop-identification tasks. The reader is reminded that /f/ was created with 0 -dB gain for each of the six formant frequencies, while /s/ and /ʃ/ were generated by adding nominal gain of 60 dB to either the sixth or the third formant, respectively. Starting with /f/ and /s/ as the beginning and end points, a continuum of 60 stimuli was created with the gain on the sixth formant changing from 0 dB to 60 dB in 1-dB steps using the Klatt (1980) speech synthesizer. Similarly, another continuum of 60 stimuli, starting with /f/ and ending with /ʃ/, was created by changing the nominal gain on the third formant from 0 dB to 60 dB in 1-dB steps. Stimulus duration was 250 ms. Cutting 10-ms segments from each of the stimuli in these continua yielded the corresponding stimuli in the /p-t/ and /p-k/ continua. Thus, the /f-s/, /f-ʃ/, /p-t/, and /p-k/ continua were generated, each containing 60 stimuli. In each case, the task of the listener was to detect the difference between the standard stimulus (/f/ or /p/) from one with additional gain on either the third (in case of /f-ʃ/ and /p-k/) or the sixth (in case of /f-s/ and /p-t/) formant. The overall rms amplitude for all stimulus waveforms within each of these continua was equalized. Stimuli were also low-pass filtered at 5000 Hz.

Preliminary pilot testing using the method of constant stimuli was done in order to identify a range of test stimuli from each continuum for final testing. The goal was to determine the range of stimulus values and step sizes needed to span a discrimination performance range of about 50% (chance) to 100% correct. Three young normal-hearing adults were tested with each of these four stimulus sets. Based on the psychometric functions obtained, 15 test stimuli were included in the /f-s/ and /p-t/ continua, with the nominal gain parameter of the sixth formant ranging from 14 to 42 dB and 20 to 48 dB, respectively, in 2-dB steps. For the /f-ʃ/ and /p-k/ continua, 20 stimuli were selected, with values for the third-formant amplitude parameter ranging from 14 to 52 dB and 20 to 58 dB, respectively, in 2-dB steps. Representative spectra from the /f-s/ and /f-ʃ/ continua are shown in Fig. 3. For each of these spectra, the arrows highlight relatively a flat frequency region (approximately 1000 Hz) and a spectral peak (at 2300 Hz in case of the /f-ʃ/ stimuli and at 4900 Hz in case of the /f-s/ stimuli). The difference between the relative amplitudes of the flat region and the spectral peak will be referred to as the peak-base difference (see Procedures below) and represents a metric of spec-

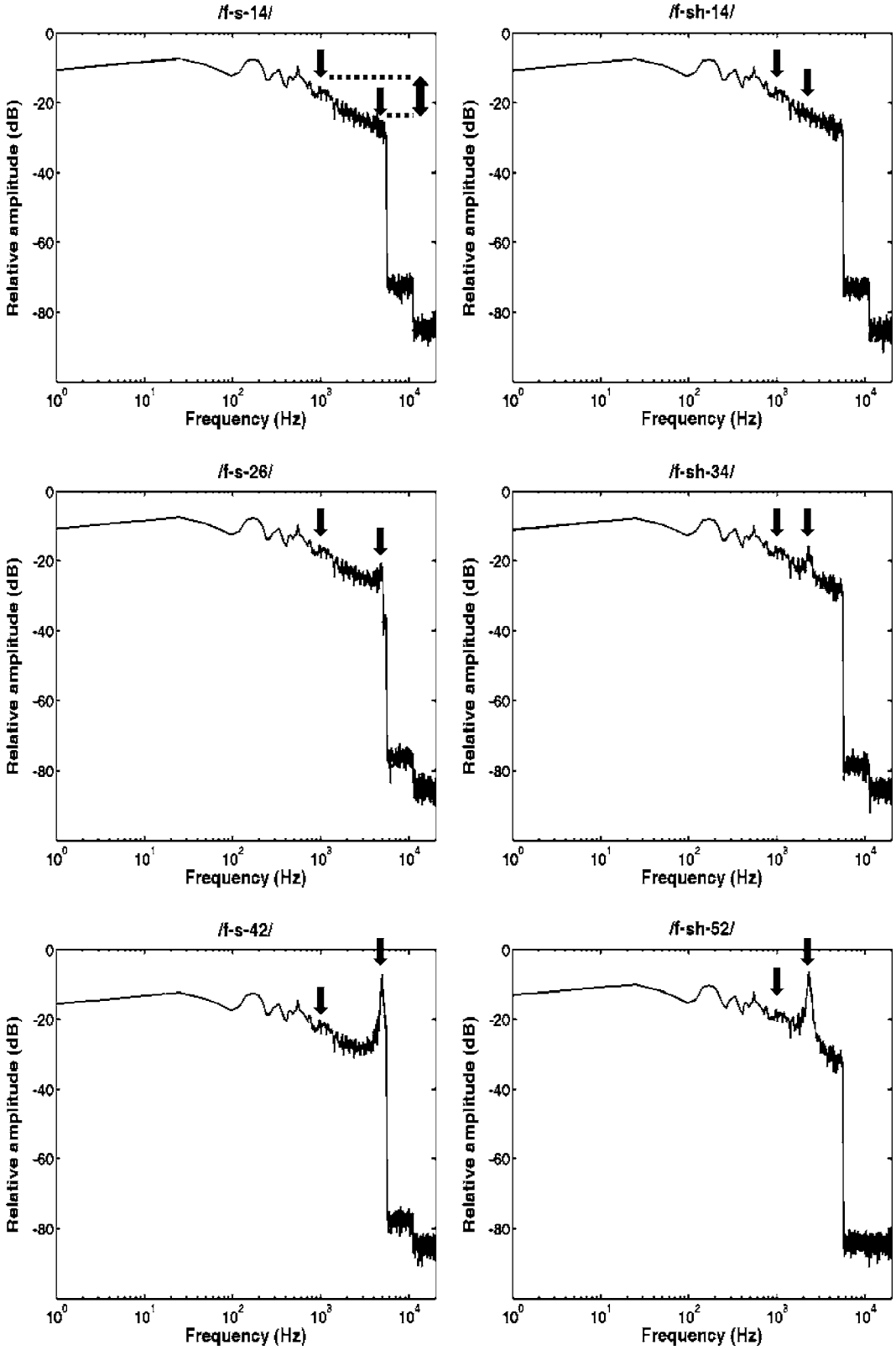


FIG. 3. Amplitude spectra for stimuli from the two endpoints and middle of the /f-s/ and /f-ʃ/ continua are shown in the three panels on the left and right, respectively. The three spectra on the left represent stimuli from the /f-s/ continuum with 14, 26, and 42 dB gain, respectively, on the sixth formant. The three spectra on the right represent stimuli from the /f-ʃ/ continuum with 14, 34, and 52 dB gain, respectively, on the third formant. In each panel, the arrows represent relatively flat regions (approximately 1000 Hz) and the spectral peaks.

tral shape for these stimuli. Note that for stimuli at the beginning of the continua (for example, /f-s-14/ and /f-ʃ-14/), the peak-base difference is negative.

b. Spectrally tilted broadband noise stimuli: These stimuli consisted of a broadband noise spectrally tilted in a symmetric manner through the logarithmic center frequency of the band. In order to generate this series of stimuli, a 10-s sample of white noise was first digitally generated at a

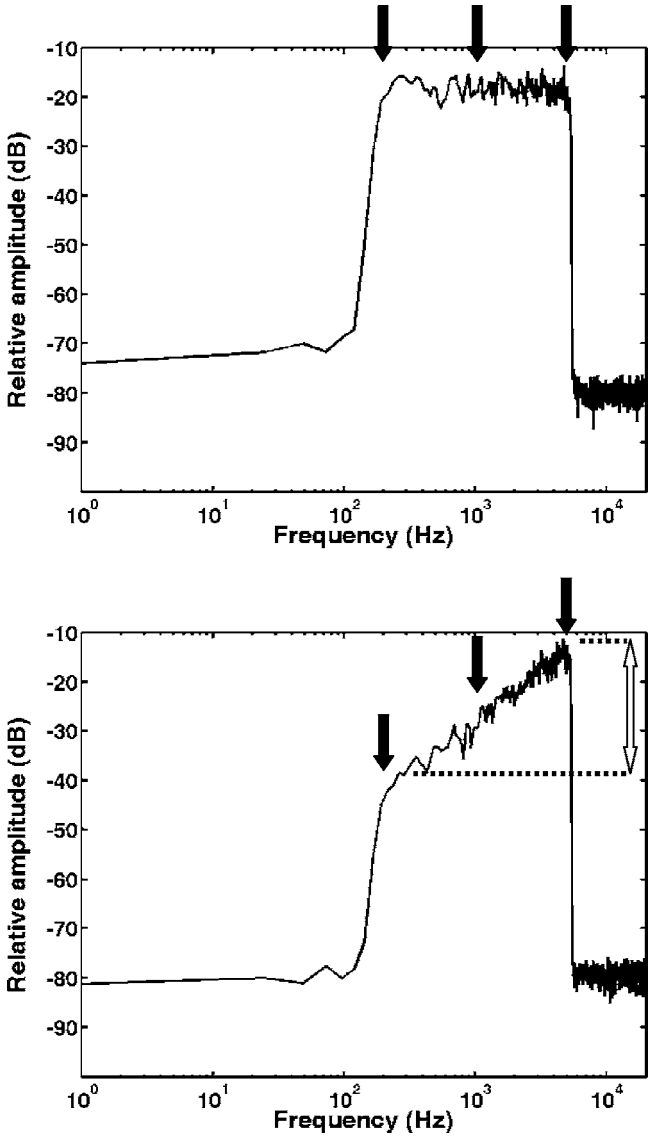


FIG. 4. Spectra with 0 and 30-dB tilt through the logarithmic center frequency are shown in the top and bottom panel, respectively. The three arrows in each panel represent the two end points (200 and 5076 Hz) and the center frequency (1007 Hz), respectively, of the broadband noise. The two-sided arrow in the bottom panel represents the amount of tilt across the two ends of the spectrum, in this case, 30 dB.

sampling rate of 44 100 Hz and bandpass filtered to give a broadband noise with lower and upper cutoffs of 200 and 5076 Hz, respectively. This stimulus had a flat spectrum, corresponding to a 0-dB tilt, as shown in the top panel of Fig. 4. A tilted spectrum was defined as one with an equivalent increment and decrement in amplitude above and below the logarithmic center frequency of the band (1007 Hz). For example, a 30-dB spectral tilt for this band of noise corresponded to a 15 dB decrement from 1007 to 200 Hz, and a 15-dB increment from 1007 to 5076 Hz. A stimulus with such a 30-dB tilt is shown in the bottom panel of Fig. 4. The direction of spectral tilt was always in the positive direction from low to high frequencies, as illustrated. Given the presence of high-frequency hearing loss in the EHI listeners, stimuli with negatively sloping tilts would have run the risk of being inaudible. This is also the direction of spectral tilt in the synthesized fricative and stop consonants (Fig. 3).

Once a series of stimuli with tilts ranging from 0 to 40 dB in 1-dB steps were generated, 250-ms and 25-ms samples were cut from each of them at zero crossings of the waveform. The final set of stimuli was comprised of 20 250-ms and 20 25-ms broadband noise stimuli with spectral tilts ranging from 0 to 34 dB. In each case, the task of the listener was to detect the difference between a flat stimulus and one with a spectral tilt across the logarithmic center frequency. These stimuli were all resampled at 48 828 Hz and digitally equalized for rms amplitude.

3. Electroacoustic calibration

In order to ensure audibility across the entire spectrum for all the stimuli, it was necessary that for every subject tested, all stimuli be above threshold at each frequency. In order to accomplish this, amplitude spectra like those in Figs. 2–4 were generated for *all* the spectrally tilted broadband-noise and speechlike stimuli. When the full set of amplitude spectra were superimposed, amplitude minima and maxima were noted across the spectrum and noise stimuli representing these maxima and minima were generated for calibration purposes. Determination of amplitude maxima and minima included consideration of the within-trial amplitude rove used in the study (0–15 dB, see Procedures). Overall output levels and one-third octave-band levels were measured using ER-3A insert earphones in a 2-cm^3 coupler for the calibration noise stimuli representing the maximum and minimum spectral amplitude envelopes. These values were then used to define: (1) the maximum amount of permissible hearing loss at each frequency for the hearing-impaired subjects, such that the stimuli representing the minimum amplitude would be above threshold at each frequency; and (2) the maximum levels presented to ensure safe presentation levels and absence of saturation. Stimuli were presented such that the calibration stimulus representing the maximum amplitude was at an overall level of 102 dB SPL for all listeners. Such a high presentation level was necessary to ensure audibility for the hearing-impaired listeners across the entire test-frequency range and it only occurred when a stimulus with maximum amplitude was presented with a +15 dB rove. The corresponding amplitude minima are presented in Fig. 1 as the "+" symbols. As noted previously, Fig. 1 also includes the hearing thresholds of the EHI listeners (filled triangles with error bars) in the present study. It is apparent that the full bandwidth of all the stimuli was at least minimally audible to all listeners from 250 through 4000 Hz.

C. Procedures

All experimental testing was completed in a laboratory with provision for independently testing multiple subjects simultaneously. The subjects were seated comfortably in a sound-treated booth that complied with ANSI (1991) standards for ambient noise suitable for threshold measurements. Subjects sat in cubicles facing a high-resolution 15 in. color computer monitor. All stimuli were presented monaurally through ER-3A insert earphones placed in both ears.

The tests were administered to all subjects in the same order. The NST was administered first, followed by identification of the three synthesized fricatives and then the three stops. Finally, discrimination was tested for the long and

short spectrally tilted stimuli, and the four synthesized speech continua in the following order: 250-ms spectrally tilted stimuli, /f-s/, /f-ʃ/, 25-ms spectrally tilted stimuli, /p-t/ and /p-k/. All tests were completely self-paced and response time measures were not collected. The entire testing was completed in 3–4 sessions of 90–120 min each. The procedural details of each test are described below in order.

In the original NST, each of the test items is presented once and responses are collected on an answer sheet containing seven to nine alternatives for each syllable. In the present study, a desktop computer was used to administer the NST. For each subtest, all of the 7–9 response alternatives for a given subtest were displayed on the computer monitor. Subjects made their responses by positioning the computer mouse on the screen and clicking their choice from the alternatives displayed. No feedback was provided on any of the trials. Unlike the original NST, each test item was presented three times, and the number of trials for each subtest varied between 21 (7 response alternatives) and 27 (9 response alternatives). The order of stimuli within each subtest was random. However, the same order of subtests was maintained across all subjects in order to minimize all sources of inter-listener variability that could arise due to factors other than the listeners' individual abilities. Specifically, the subtests were administered in the following order: "Ca1," "cu," "Ca2," "ic," "iC," "ca," "ac," "aC," "uC," "uc," and "ci." Performance was scored in terms of percent correct on each subtest and for the entire NST, and these scores were converted to rationalized arcsine units (rau) to stabilize the error variance (Studebaker, 1985).

Identification of the synthesized fricative and stop stimuli was tested separately. In each case, a closed-set response format with three alternatives (/f/, /s/, and /ʃ/ in case of fricative-identification and /p/, /t/, and /k/ in case of stop-identification) was used, and the task was completed in four blocks of 90 trials (3 stimuli $\times$ 30 repetitions). The first block of 90 trials involved feedback on every trial in order to provide some training to the subjects in the labeling of these synthesized stimuli. Feedback consisted of a message indicating that the subject's response was either "correct" or "incorrect" and, if incorrect, the correct response alternative. The remaining three test blocks did not include feedback, and only the scores from these blocks were considered for analyses. Performance was scored as percent correct and transformed into rau units.

The presentation of stimuli for the fricative-and stop-identification tasks was similar to that for the NST. In each case, all three alternatives were displayed on the computer monitor, and subjects made their responses using a computer mouse.

Discrimination thresholds were obtained separately for each of the four stimulus sets (/f-s/, /f-ʃ/, /p-t/ and /p-k/). For each stimulus set, the 70.7 percent-correct point on the psychometric function performance was established using a three-interval (standard, two-comparison), two-alternative, forced choice procedure with two-down, one-up adaptive tracking (Levitt, 1971) with a 400-ms interstimulus interval. The standard stimulus (either /f/ or /p/) was always presented in the first interval, with a comparison stimulus in each of the remaining two intervals. For the /f-s/ and /f-ʃ/ stimulus sets, /f/ was always the standard, with one of the other stimuli in the continua serving as one of the comparison stimuli. Similarly, for the /p-t/ and /p-k/ stimuli, /p/ was always the standard, with one of the other stimuli in the continua as one of the comparison stimuli. On each trial, the listener indicated the comparison stimulus that was perceived to be different from the standard. For each of the six discrimination conditions, six blocks of 50 trials were presented with feedback on every trial. Although all stimuli were equalized in overall rms amplitude, an intensity rove of 0 to +15 dB was implemented across the intervals in a trial in order to further reduce the usefulness of overall amplitude as a cue for discrimination. It is to be noted that this amount of amplitude rove was insufficient (Green, 1988) to ensure that listeners would base their responses on the spectral shape of the stimuli rather than overall loudness cues. However, it was impossible to use higher levels of rove in light of the limited dynamic range of the hearing impaired listeners in this study. Independent threshold estimates were collected from each of the six blocks by starting every block with the stimulus with the largest difference from the standard.

A methodology similar to that for the synthesized speech tokens was employed for discrimination for the spectrally tilted broadband noise stimuli as well. In this case, the standard was always the flat-spectrum noise, while the comparison stimulus had some degree of tilt. For the 250-ms stimulus, 15 spectral tilt values were used beginning with 2 dB of tilt and ending with 30 dB of tilt in 2-dB steps. For the 25-ms stimulus, 15 spectral tilt values were used beginning with 6 dB of tilt and ending with 34 dB of tilt in 2-dB steps. A random 0 to +15 dB within-trial amplitude rove was implemented to ensure that subjects made use of the relative amplitude differences across frequency rather than differences in overall amplitude.

III. RESULTS

A. Between-group comparisons

1. Speech identification

a. Scores on the NST: The mean rau-transformed (Studebaker, 1985) percent-correct scores, and associated standard errors of the mean for the two groups of listeners (YNH and EHI) on the 11 subtests of the NST are shown in Fig. 5. A series of independent samples t-tests on the transformed scores revealed that the YNH listeners performed significantly better ($p < 0.05$) than the EHI listeners on 6 of the 11 subtests: "cu," "Ca2," "ic," "iC," "ca," and "aC." Here, and throughout the paper, unless otherwise noted, p values for t-tests were adjusted for multiple comparisons. This was the case despite ensuring minimum audibility through at least 4000 Hz for the EHI listeners, and reducing the involvement of cognitive factors using nonsense syllables and a closed-set response format. It should be noted that the group differences between the YNH and EHI listeners on these subtests may be due to the signals not being *equally* audible (i.e., equivalent sensation levels) for the two groups of listeners. However, age-related changes in NST scores have been reported even in normal-hearing adults (Gelfand *et al.*, 1986), and the poorer performance of EHI listeners in the present study was not entirely unexpected.

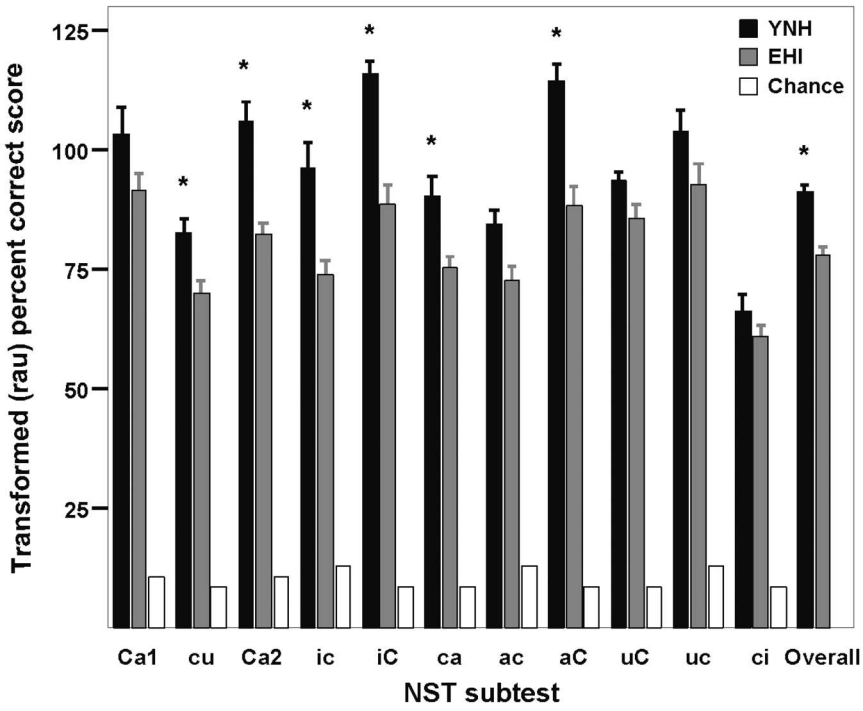


FIG. 5. Mean transformed (rau) percent-correct scores and standard error of the mean for the YNH (black bars) and EHI (grey bars) subjects on the overall NST and its subtests. Significant group differences in performance are indicated by $^{*}(p<0.01)$. Chance performance is indicated by the white bars.

An overall NST score, which represents the percentage of the 273 (91 stimuli $\times$ 3 repetitions) tokens identified correctly, was also computed. This score appears in Fig. 5 as well (far right). Overall, the EHI listeners identified 78% of the nonsense syllables correctly, which was significantly ($p<0.01$) less than the 91% correctly identified by the YNH listeners. The poorer performance of the YNH listeners in the present study, when compared to nearly 100% correct performance reported by Dubno and Levitt (1981), may be related to the fact that all the NST tokens were spectrally shaped and presented at a higher sound level to ensure audibility for the EHI listeners. Bandpass filtering (200–5000 Hz) of the stimuli may have also played a role.

b. Identification of the synthesized fricative and stop tokens: rau-transformed percent-correct scores from the last three blocks (without feedback) for the fricative- and stop-identification tasks were subjected to a repeated-measures analysis of variance (ANOVA) to determine if performance changed significantly over the three blocks and in a systematic manner. The analyses did not indicate a significant effect of block number for the fricatives $[F(2,27)=0.595,\ p>0.05]$ or for the stops $[F(2,27)=2.719,\ p>0.05]$. Hence, the data from these three blocks were reduced to a single mean transformed score for each of the two stimulus sets and used for all further analyses. Mean rau-transformed scores on the fricative- and stop-identification tasks for the two groups of listeners are presented in Fig. 6. Independent-sample t-tests indicated that the YNH subjects had significantly higher ($p<0.05$) identification scores than the EHI subjects for the fricative stimuli only. Note, however, that both groups of subjects had considerable difficulty identifying the synthesized stop consonants.

2. Discrimination thresholds

Discrimination thresholds are expressed either as peak-base differences for the speechlike stimuli (Fig. 3) or as spectral tilt values for the broadband noise stimuli (Fig. 4). Repeated-measures ANOVAs were performed on the discrimination thresholds from each of the six blocks of trials for each stimulus set and revealed a significant change in thresholds over the first two blocks for two of the stimulus sets (/f-s/ continuum and the 250-ms nonspeech stimuli). For this reason, data from the first two blocks for all six stimulus sets were discarded, and mean discrimination thresholds were based on the final four blocks of 50 trials each (200 trials total).

The mean discrimination thresholds for the two groups

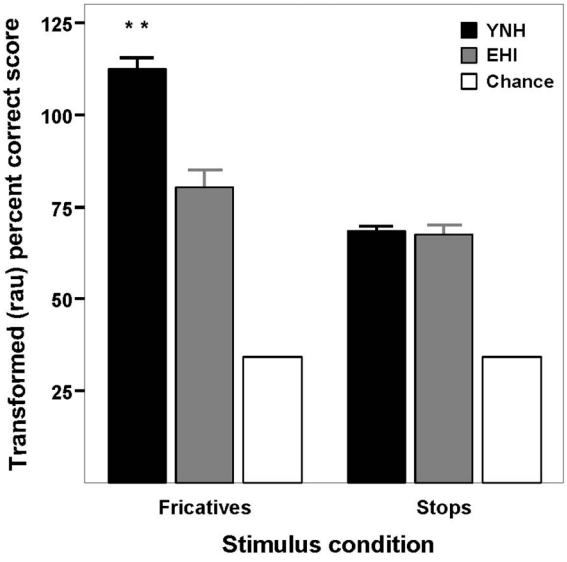


FIG. 6. Mean transformed percent-correct identification scores and standard error of the mean (vertical lines) for the YNH (black bars) and EHI (grey bars) listeners for the synthesized fricative and stop stimuli. Significant group differences are indicated by $^{**}(p<0.01)$. Chance performance is indicated by the white bars.

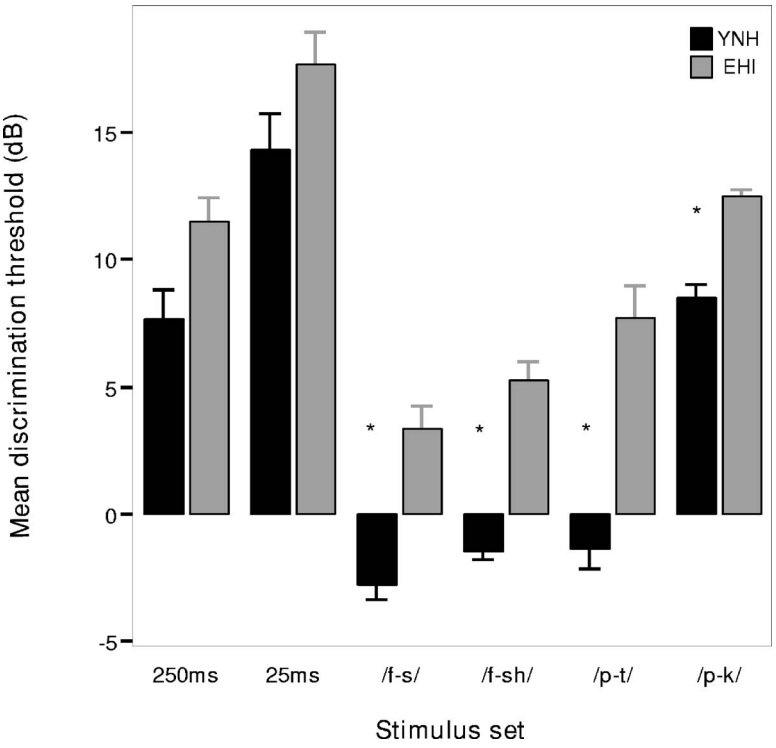


FIG. 7. Mean discrimination thresholds and standard error of the mean for the YNH (black bars) and EHI (grey bars) listeners for the six stimulus conditions. Significant group differences are indicated by $^{*}(p<0.01)$.

of listeners for the six stimulus conditions are presented in Fig. 7. Independent-samples t-tests were performed to determine group differences in discrimination thresholds for each of the six stimulus sets. The results revealed that thresholds for the YNH listeners were significantly lower than thresholds for the EHI listeners for all speechlike stimulus sets. These results are in agreement with studies on spectral-shape discrimination in elderly hearing-impaired listeners (Summers and Leek, 1994; Lentz and Leek, 2002).

B. Within-group analyses

The following sections describe stimulus effects within each group of subjects. Results from the YNH listeners are presented first, followed by results for the EHI listeners. In addition, correlational and regression analyses are presented for the EHI subjects.

1. YNH listeners

A paired-sample t-test on the two sets of synthesized speech stimuli revealed a significant effect of stimulus condition on identification performance $[t(9)=16.51,\ p<0.01]$. The YNH listeners were significantly better at identifying the synthesized fricatives than the stops. This result was not unexpected because of the very short duration and the lack of formant transitions in the synthetic stop stimuli. The YNH listeners exhibited confusions between /p/ and /k/ stimuli, a result that is similar to that of Miller and Nicely (1955) who measured open-set stop identification in conditions of varying background noise and stimulus bandwidth. However, the identification scores of the present group of YNH listeners in *quiet* was poorer than the listeners from Miller and Nicely (1955) in noise. These results suggest that in the present study, the derivation of stop stimuli by shortening the duration of the fricative stimuli did not result in good exemplars for /p/ and /k/ in either group of listeners.

A regression analysis was performed to predict the NST speech-identification scores from several independent variables. The overall NST score (in rau) was the dependent variable, with age, rau-transformed percent-correct scores for identification of fricatives and stops, discrimination thresholds for the 250-ms and 25-ms nonspeech stimuli, and discrimination thresholds for the /f-s/, /f-ʃ/, /p-t/, and /p-k/ stimulus sets, as the independent or predictor variables. No significant predictive relationship evolved from the regression analysis. This can be attributed primarily to the homogeneity in the speech-identification scores of the YNH listeners.

2. EHI listeners

The correlations between thresholds on the six discrimination stimulus sets are shown in Table II. In general, weak to moderate positive correlations were observed, but only about half were statistically significant. Thresholds for the two spectrally-tilted broadband-noise stimuli were positively correlated, indicating that lower thresholds for the 250-ms stimuli were associated with lower thresholds for the 25-ms stimuli. Positive correlations were observed between thresholds for the /f-s/ and /p-t/ stimulus sets and for the /f-ʃ/ and /p-k/ stimulus sets. These results are not surprising as these two sets of stimuli are spectrally similar and differ only in duration. Table II also includes correlations between fricative- and stop-identification scores, thresholds for the six discrimination stimulus sets, and the overall NST score. It can be observed that thresholds for the /f-ʃ/, /p-t/, and /p-k/ stimulus sets were negatively correlated ($p<0.05$) with the fricative-identification score, while thresholds for the 25-ms spectrally tilted broadband-noise, /p-t/, and /p-k/ stimulus

TABLE II. Pearson correlations (r) between thresholds for the six discrimination stimulus sets, and fricative- and stop-identification scores for the EHI listeners. Significant correlations are marked with $^{*}(p<0.05)$ and $^{**}(p<0.01)$.

Stimulus	250-ms	25-ms	/f-s/	/f-ʃ/	/p-t/	/p-k/	Fric-Ident	Stop-Ident	Overall NST
250-ms	1	0.49*	0.28	0.37	0.23	0.17	−0.19	−0.24	−0.21
25-ms		1	0.57**	0.40	0.60**	0.33	−0.44	−0.51*	−0.48*
/f-s/			1	0.33	0.66**	0.00	−0.33	−0.44	−0.61**
/f-ʃ/				1	0.46*	0.59**	−0.55*	−0.41	−0.37
/p-t/					1	0.41	−0.54*	−0.51*	−0.35
/p-k/						1	−0.49*	−0.52*	−0.18
Fric-Ident							1	0.53*	0.35
Stop-Ident								1	0.60**
Overall NST									1

sets were negatively correlated ($p<0.05$) with the stop-identification score. Discrimination thresholds for the 25-ms and /f-s/ stimulus conditions were negatively correlated with the overall NST score. These results indicate that, in general, the smaller (better) the discrimination threshold, the higher (better) the identification scores. Finally, identification for the synthetic stops was positively correlated with the overall NST score, indicating that better the identification of the synthetic stops, better the identification performance on the overall NST.

This study was primarily aimed at investigating individual differences in speech-identification and spectral-shape discrimination among elderly hearing-impaired listeners. With this in mind, a linear stepwise regression analysis (probability of F to enter $\leqslant 0.05$; probability of F to remove $\geqslant 0.10$) was then performed on the data from the EHI listeners in order to reveal predictive relationships between the variables tested in this study. The regression analysis involved the score for the overall NST (in rau) as the dependent variable, with age, rau-transformed percent correct scores for identification of fricatives and stops, discrimination thresholds for the 250-ms and 25-ms nonspeech stimuli, discrimination thresholds for the /f-s/, /f-ʃ/, /p-t/, and /p-k/ continua, pure-tone average at 500, 1000, and 2000 Hz (PTA), and high-frequency pure-tone average at 1000, 2000, and 4000 Hz (HFPTA), as the independent or predictor variables. Two variables, the discrimination threshold (dB) for the /f-s/ condition, and the identification score for stops, emerged as significant predictor variables for the overall NST score. An adjusted R-square of 0.34 ($p<0.01$) was found between the overall NST score and the discrimination threshold for the /f-s/ stimulus set, which improved to 0.45 ($p<0.01$) when the identification score for stops was added. The final regression equation for these variables was as follows: overall NST score (in rau)$=0.287\ x$ stop identification $-0.843\ x$ /f-s/ discrimination threshold $+61.47$. The overall NST score was positively associated with the mean stop identification score ($r=0.60$, $p<0.01$) and negatively associated ($r=-0.61$, $p<0.01$) with the mean /f-s/ discrimination threshold. Regarding the latter association, performance on the NST improved as discrimination thresholds decreased (got better). The measured overall NST score and the predicted overall NST score (from the regression equation) was positively correlated ($r=0.70$, $p<0.01$).

The subtests of the NST consisted of consonants such as nasals, glides, affricates, and liquids in addition to fricatives and stops. The predictor variables used in the regression, on the other hand, are based on fricative and stop consonants. In order to determine if the predictability of the NST score would improve if the NST score were only based on fricatives and stops, a fricative-stop NST score was calculated by evaluating only the fricative and stop consonants from each of the subtests. A regression analyses was then conducted with this fricative-stop NST score and the same predictor variables as before, to determine if removing other types of consonants would improve the contribution of the predictor variables. The results showed that spectral-shape discrimination threshold (dB) for the /f-s/ condition was the only significant predictor variable for this fricative-stop NST score. The adjusted R-square value was 0.34 ($p<0.01$) for this regression, identical to that observed for the overall NST score and this same predictor variable. The lack of contribution of the two stop continua to the fricative-stop NST score also supports the conclusion that the derivation of the stop stimuli in the present study by shortening the corresponding fricative stimuli did not result in the best exemplars for the stop stimuli.

IV. GENERAL DISCUSSION

Results from the regression analysis for scores within the EHI listeners revealed that when all stimuli were at least minimally audible across the full stimulus bandwidth (200–5000 Hz), about 36% of the variance in speech-identification scores could be attributed to spectral-shape discrimination abilities. This relationship, although significant, is, at best, moderate. Also, the stimuli in the present study were tailored such that the tokens used for spectral-shape discrimination were similar to those used for speech-identification. Nevertheless, the results are interesting because they differ from the findings of previous studies that did not find any contribution of auditory processing abilities to speech-understanding in quiet. One possible explanation

for the differences in findings could be that most previous studies have involved basic auditory processing skills, such as frequency and temporal resolution for simple and complex sounds, whereas the present study involved spectral-shape discrimination, arguably a more "complex" auditory processing task. In light of this, the further study of complex auditory processing abilities, such as spectral-shape discrimination, in elderly hearing-impaired listeners seems warranted.

It is important to note that these results were obtained while ensuring audibility of all stimuli for all listeners. Perhaps as important, however, is the finding that neither age, nor hearing loss, entered into any of the regression solutions discussed above. These results are somewhat surprising as many studies have associated these two variables with unaided (Humes and Roberts, 1990; Humes and Christopherson, 1991; Gordon-Salant and Fitzgibbons, 1997; Humes *et al.* 1994), and aided (Humes, 2002) speech-identification and auditory processing in elderly listeners. For example, Humes (2002) found that the primary factor contributing to performance on unaided and aided speech-recognition tasks in elderly hearing-impaired listeners was the amount of hearing loss. The NST stimuli in the present study were shaped in a manner similar to well-fit amplification using a two-channel hearing aid. Hence, the finding that the amount of hearing loss was not related to the NST score can be attributed to the fact that all the NST stimuli were at least minimally audible for all listeners. For the rest of the stimuli, audibility was ensured for the hearing-impaired listeners by presenting the full stimulus bandwidth *at least* 1–2 dB above corresponding pure-tone thresholds. This minimum criterion, however, was the worst-case scenario and was only true for the lowest amplitude spectrum and at the lowest rove intensity. The majority of stimuli presented far exceeded this minimum. Nonetheless, this criterion is far from optimal audibility, and hence, a contribution of hearing loss to speech-identification scores for these stimuli would not have been entirely unexpected. All the hearing-impaired listeners in the study had hearing loss in the mild to moderate range. The relatively mild amounts of hearing loss in the elderly hearing-impaired subjects might explain the lack of contribution of hearing loss to the regression solutions for these stimuli.

As mentioned above, age did not contribute to the variance in speech-identification for natural and synthetic speech tokens. This might be explained by the fact that all the speech-identification tasks in the study involved simple, nonsense-syllable stimuli presented in relatively low-uncertainty conditions and using closed-set response formats. This may have minimized the contribution of age related cognitive factors that might be associated with performance on the speech-identification tasks. However, it should be kept in mind that there *were* significant group differences between the YNH and EHI groups, and a portion of these differences could be due to associated differences in age. *Within the EHI group*, on the other hand, individual differences in age were not associated with individual differences in speech-identification performance.

It is also interesting to note that spectral-shape discrimination thresholds for the two broadband-noise stimuli did not contribute to any of the regression solutions for speech-identification. On the other hand, discrimination thresholds for spectral-shape changes localized to small frequency regions were moderately related to speech-identification scores. These results suggest that listeners' ability to discriminate spectral-shape changes across broad frequency regions may not be related to identification of stimuli with spectral-shape changes localized to narrow frequency regions. These results may simply be related to the long-term experience that the listeners had with speech stimuli in general.

The finding that the listeners performed worse on the spectrally-tilted broadband-noise stimuli than on the speechlike stimuli may indicate that they were better at discriminating spectral-shape differences that are localized to narrow frequency regions. This finding is different from that of profile analysis studies and may be related to the nonuniform nature of the stimuli used here. Additional research is required to address this difference in performance across stimuli.

It should be noted that these results were obtained in quiet conditions. It is known that the speech-understanding difficulties of the elderly hearing-impaired are compounded in the presence of background noise, and that auditory processing abilities, such as temporal processing, have been found to contribute to speech-understanding in noise. Thus, it is important to extend the present work to an examination of speech-identification in noise.

V. SUMMARY

The following is a summary of the main findings of this study:

(a) The YNH listeners performed significantly better than the EHI listeners on several of the identification tasks (including six of the subtests of the NST, overall NST, and identification of the synthesized fricative consonants). The YNH listeners also performed better than the EHI on all four discrimination tasks involving speechlike stimuli, but no significant group differences were obtained for the two broadband noise stimulus sets.
(b) Within each group of listeners, significantly higher scores were obtained for the fricative identification task than on the stop-identification task. For both groups of listeners, discrimination thresholds for the speechlike tasks were better (smaller) than those for the broadband noise stimuli.
(c) When all the stimuli were at least minimally audible, spectral-shape discrimination abilities of elderly hearing-impaired listeners had a moderate and significant association with the speech-identification scores in quiet.
(d) Within the EHI listeners, there were no effects of hearing loss or age on any of the measures.

ACKNOWLEDGMENTS

The authors wish to thank Jennifer Lentz and Charles Watson for their contributions to the design of the study, and

Dana Wilson for her assistance with participant recruitment. The research presented in this manuscript was supported, in part, by an NIH Grant No. R01 AG08293-12 awarded to L.E.H.

ANSI (**1991**). ANSI S3.1-1991, *Maximum permissible ambient noise levels for audiometric test rooms*. American National Standards Institute, New York.

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

CHABA. (**1988**). "Speech understanding and aging. Working Group on Speech Understanding and Aging. Committee on Hearing, Bioacoustics, and Biomechanics, Commission on Behavioral and Social Sciences and Education, National Research Council," J. Acoust. Soc. Am. **83**, 859–895.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.

Flynn, M. C., Dowell, R. C., and Clark, G. M. (**1998**). "Aided speech recognition abilities of adults with a severe or severe-to-profound hearing loss," J. Speech Lang. Hear. Res. **41**, 285–299.

Gelfand, S. A., Piper, N., and Silman, S. (**1986**). "Consonant recognition in quiet and in noise with aging among normal hearing listeners," J. Acoust. Soc. Am. **80**, 1589–1598.

Gordon-Salant, S. (**1987**). "Age-related differences in speech recognition performance as a function of test format and paradigm," Ear Hear. **8**, 277–282.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1995a**). "Comparing recognition of distorted speech using an equivalent signal-to-noise ratio index," J. Speech Hear. Res. **38**, 706–713.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1995b**). "Recognition of multiply degraded speech by young and elderly listeners," J. Speech Hear. Res. **38**, 1150–1156.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford University Press, London).

Heinz, J. M., and Stevens, K. N. (**1961**). "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. **33**, 589–596.

Humes, L. E. (**2002**). "Factors underlying the speech-recognition performance of elderly hearing-aid wearers," J. Acoust. Soc. Am. **112**, 1112–1132.

Humes, L. E., and Christopherson, L. (**1991**). "Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits," J. Speech Hear. Res. **34**, 686–693.

Humes, L. E., and Roberts, L. (**1990**). "Speech-recognition difficulties of the hearing-impaired elderly: the contributions of audibility," J. Speech Hear. Res. **33**, 726–735.

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (**1994**). "Factors associated with individual differences in clinical measures of speech recognition among the elderly," J. Speech Hear. Res. **37**, 465–474.

Jerger, J., Jerger, S., Oliver, T., and Pirozzolo, F. (**1989**). "Speech understanding in the elderly," Ear Hear. **10**, 79–89.

Johnson, K. (**1987**). "KLSYN: A Formant Synthesis Program," Retrieved July 1, 2003, from http://www.indiana.edu/~acoustic/klsynman.html

Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (**1983**). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **73**, 1779–1793.

Klatt, D. L. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Lentz, J. J., and Leek, M. R. (**2002**). "Decision strategies of hearing-impaired listeners in spectral shape discrimination," J. Acoust. Soc. Am. **111**, 1389–1398.

Lentz, J. J., and Leek, M. R. (**2003**). "Spectral shape discrimination by hearing-impaired and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 1604–1616.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Levitt, H., and Resnick, S. B. (**1978**). "Speech reception by the hearing-impaired: Methods of testing and the development of new tests," Scand. Audiol. Suppl. **6**, 107–130.

Marshall, L. (**1981**). "Auditory processing in aging listeners," J. Speech Hear Disord. **46**, 226–240.

Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Moore, B. C. (**2003**). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," Otol. Neurotol. **24**, 243–254.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Skafte, M. D. (**2000**). "The 1999 hearing instrument market—the dispenser's perspective," Hear. Res. **7**(6), 8–40.

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Summers, V., and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.

van Rooij, J. C., and Plomp, R. (**1990a**). "Auditive and cognitive factors in speech perception by elderly listeners," Acta Oto-Laryngol., Suppl. **476**, 177–181.

van Rooij, J. C., and Plomp, R. (**1990b**). "Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses," J. Acoust. Soc. Am. **88**, 2611–2624.

van Rooij, J. C., and Plomp, R. (**1992**). "Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion," J. Acoust. Soc. Am. **91**, 1028–1033.

van Rooij, J. C., Plomp, R., and Orlebeke, J. F. (**1989**). "Auditive and cognitive factors in speech perception by elderly listeners. I. Development of test battery," J. Acoust. Soc. Am. **86**, 1294–1309.

Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry

M. Kathleen Pichora-Fuller[a] and Bruce A. Schneider
Department of Psychology, University of Toronto, 3359 Mississauga Road N, Mississauga, Ontario, Canada L5L 1C6

Nancy J. Benson
The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8

Stanley J. Hamstra
Department of Surgery, University of Toronto, 100 College Street, Toronto, Ontario, Canada M5G 1L5

Edward Storzer
McNeill Audiology, 116-2187 Oak Bay Avenue, Victoria, British Columbia, Canada V8R 1G1

(Received 3 December 2004; accepted 11 November 2005)

Gap detection thresholds for speech and analogous nonspeech stimuli were determined in younger and older adults with clinically normal hearing in the speech range. Gap detection thresholds were larger for older than for younger listeners in all conditions, with the size of the age difference increasing with stimulus complexity. For both ages, gap detection thresholds were far smaller when the markers before and after the gap were the same (spectrally symmetrical) compared to when they were different (spectrally asymmetrical) for both speech and nonspeech stimuli. Moreover, gap detection thresholds were smaller for nonspeech than for speech stimuli when the markers were spectrally symmetrical but the opposite was observed when the markers were spectrally asymmetrical. This pattern of results may reflect the benefit of activating well-learned gap-dependent phonemic contrasts. The stimulus-dependent age effects were interpreted as reflecting the differential effects of age-dependent losses in temporal processing ability on within- and between-channel gap detection. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2149837]

PACS number(s): 43.71.Lz, 43.66.Sr [JHG] Pages: 1143–1155

I. INTRODUCTION

It is widely known that older adults experience greater difficulty than do younger adults when listening to speech in noise, even though they may have relatively little difficulty in quiet listening situations (e.g., CHABA, 1988; Pichora-Fuller, Schneider and Daneman, 1995; Frisina and Frisina, 1997). For more than two decades, one hypothesis has been that age-related declines in auditory temporal processing contribute to these age-related difficulties in listening to complex signals in complex environments (for reviews see Bergman, 1980; Divenyi and Simon, 1999; Schneider and Pichora-Fuller, 2000, 2001; Pichora-Fuller and Souza, 2003). However, the specific relationships between aspects of auditory temporal processing and speech perception in aging are still not fully understood.

Gap detection threshold, the smallest gap that a listener can detect in a stimulus, is a commonly investigated psychoacoustic measure of auditory temporal processing. Over the last decade, mounting evidence has suggested that many older adults do not detect gaps until they are significantly longer than those that can be detected by younger adults and that these differences are not simply attributable to audiometric threshold elevations (e.g., Schneider, Pichora-Fuller, Kowalchuk, and Lamb, 1994; Snell, 1997; Snell and Frisina, 2000). Studies not concerned with the effect of age on gap detection have also established that gap detection threshold is not correlated with degree of pure-tone hearing loss (Florentine and Buus, 1984) or with measures of the frequency selectivity of auditory filters (Grose, Eddins, and Hall, 1989; Eddins, Hall, and Grose, 1992; Moore, Peters, and Glasberg, 1993). We have learned that age-related differences are more pronounced when the markers surrounding the gap are shorter than 10 ms (Schneider and Hamstra, 1999) and when the location of the gap is near to the onset or offset of the signal (He, Horwitz, Dubno, and Mills, 1999), but that age-related differences are not related to other characteristics of the markers such as envelope shape and intensity so long as the marker levels are not near absolute threshold (Schneider, Speranza, and Pichora-Fuller, 1998). There is some suggestion that age-related differences may vary with the frequency composition of the markers in that age was correlated with gap threshold for younger adults when noise markers had a 1-kHz low-pass frequency cutoff but not for a 6-kHz cutoff (Snell and Frisina, 2000). Furthermore, age effects have been observed for gap discrimination thresholds measured when the frequency of the leading marker was fixed and the frequency of the lagging marker was varied (Lister, Koehnke, and Besing, 2000; Lister, Besing, and Koehnke, 2002; see

[a]Author to whom correspondence should be addressed. Electronic mail: kpfuller@utm.utoronto.ca; Fax: 905-569-4326; Phone: 905-828-3865.

also Grose, Hall, and Buss, 2001). Similarly, age-related declines were found on gap discrimination threshold in synthetic speech stimuli, with poorer performance on spectrally dynamic markers than on more spectrally stable markers (Lister and Tarver, 2004). Of the properties of the marker that have been related to age differences in gap detection and discrimination, the duration and the spectral symmetry of the markers are both factors that are likely to influence the usability of gap cues for speech perception.

Gaps in speech signals can cue phonetically important distinctions; for example, a gap is one cue to the presence of the stop consonant phoneme /p/ in the word "spoon," in contrast to the word "soon" in which there is no gap. To the extent that psychoacoustic gap detection thresholds measure a listener's ability to perceive the smallest detectable gap in a sound signal and the detection of such gaps is relevant to speech perception, we might expect to find a strong relationship between measures of gap detection threshold and speech perception ability. Some previous studies have found significant correlations between measures of gap detection in nonspeech stimuli and word identification in noise (e.g., Tyler, Summerfield, Wood, and Fernandes, 1982; Snell, Mapes, Hickman, and Frisina, 2002) or reverberation (Gordon-Salant and Fitzgibbons, 1993). However, other studies have not found significant correlations between gap detection thresholds and speech measures such as syllable identification or discrimination in quiet based on voice onset time contrasts (e.g., Strouse, Ashmead, Ohde, and Grantham, 1998). The degree of cochlear hearing loss has been related to declines in psychoacoustic difference limens, including gap detection in noise and in speechlike composite signals, but performance on the psychoacoustic measures did not explain differences in response patterns to speech stimuli (Nelson, Nittrouer and Norton, 1995). Differences in the specific stimuli used to measure gap detection thresholds and/or in the specific speech stimuli used to measure speech perception could possibly account for the inconsistency in these findings concerning the relationship between psychoacoustic measures of auditory gap detection ability and the functional use of temporal cues in speech perception.

It is important to investigate how the effectiveness of a gap as a cue to stop consonant identification may depend on marker duration because (a) age-related differences in gap detection thresholds are more pronounced when the duration of the markers is short (e.g., Schneider and Hamstra, 1999), (b) age-related differences in speech understanding are more pronounced in fast speech than in slow speech (e.g., Gordon-Salant and Fitzgibbons, 2004; Wingfield, 1996), and (c) gaps cueing phoneme contrasts are more difficult to detect for older listeners (Lister and Tarver, 2004; Price and Simon, 1984; Ohde and Abou-Khalil, 2001). To investigate the relationship between gap detection and speech perception, we conducted a study in which younger and older adults detected gaps in 2-kHz tone-pips with varying marker durations and also distinguished word pairs differing in presence or absence of a stop consonant (e.g., *spoon* vs *soon, split* vs *slit, catch* vs *cash, ditch* vs *dish*) spoken at fast and slow rates in quiet and in noise (Haubert and Pichora-Fuller, 1999). Compared to their younger counterparts, older adults could not detect gaps until they were longer in both the 2-kHz tone-pips and in the speech stimuli. In addition to these age-related differences in both nonspeech and speech gap detection measures, psychoacoustic gap detection thresholds and gap thresholds required for reliable word pair discrimination correlated significantly, at least for the older group in the fast speech rate conditions in noise. Nevertheless, it remained difficult to explain how psychoacoustic gap detection thresholds for 2-kHz tone-pips could be related to the minimum size of gap required for word discrimination because the mean nonspeech gap detection thresholds never exceeded 8 ms but the mean size of gap required to cue phoneme identification in the word-pair discrimination task was always at least twice as large (Haubert and Pichora-Fuller, 1999). Of course, the speech signal surrounding the stop consonant gap was more spectrally complex than the 2 -kHz tone used in the gap detection task. In addition, the leading and lagging speech markers were spectrally asymmetrical, unlike the identical 2-kHz markers surrounding the gap in the nonspeech gap detection test.

When nonspeech markers more similar to speech markers are used, the size of the gap detection threshold is larger (near 30 ms) and approximates the size of the gap that serves listeners in making phonemic contrasts (e.g., Formby, Barker, Abby, and Raney, 1993; Phillips, Taylor, Hall, Carr, and Mossop, 1997). For phonemically significant gaps in speech, the spectral content of the speech markers would span multiple frequency bands and the spectral content of the leading and lagging markers could be similar or dissimilar. It is known that gap detection thresholds are affected by the spectral content of the leading and lagging markers. If the spectral content of the leading and lagging markers is identical, gap detection thresholds may be smaller when the markers are composed of multiple frequency bands compared to when they are composed of a single band, and this improvement is observed independently of the degree of frequency separation of the bands in the markers (e.g., Hall, Grose, and Joy, 1996; Phillips and Hall, 2000). Furthermore, concurrent "distracting" energy in flanking bands reduces the detection of a gap in a single band when the bandwidth is narrow, especially if the flanking band is comodulated or gated synchronously with the target band (Grose and Hall, 1993). However, the effect of concurrent bands is negligible when the nonoverlapping target and distractor noise bands are wider (Phillips and Hall, 2000) or when they stimulate different electrodes of a cochlear implant (van Wieringen and Wouters, 1999). In contrast, for spectrally different leading and lagging markers, gap detection thresholds are three to ten times larger than thresholds obtained with spectrally equivalent, symmetrical gap markers (Phillips *et al.*, 1997). The increase in gap detection threshold as a function of the frequency separation of the leading and lagging markers ranges from 5 ms when there is no separation to 90 ms when there is a half octave separation (Formby and Forrest, 1991). Similar effects of frequency separation of the markers have been found for dichotic presentations with the leading marker presented to the left ear and the lagging marker presented to the right ear (Formby, Gerber, Sherlock, and Magder, 1998). Larger gaps thresholds have also been found

TABLE I. Audiometric profile of younger and older groups.

	Younger group (N=16)				Older group (N=16)			
	Left ear		Right ear		Left ear		Right ear	
	Mean threshold (dB HL)	SD	Mean threshold (dB HL)	SD	Mean threshold (dB HL)	SD	Mean threshold (dB HL)	SD
250 Hz	2	5.8	3	7.1	13	6.8	11	5.2
500 Hz	0	6.3	0	4.6	11	6.7	11	5.2
1000 Hz	−2	5.1	0	5.5	14	7.2	12	5.4
2000 Hz	−2	5.8	−1	5.6	15	6.7	14	7.7
3000 Hz	−1	5.5	0	4.8	16	6.4	16	7.4
4000 Hz	1	6.4	0	4.3	25	14.4	23	12.1
6000 Hz	8	9.1	5	6.2	40	21.5	36	14.2
8000 Hz	3	8.9	3	5.2	49	17.7	43	16.1

when the bandwidth of the leading and lagging markers differ, even when the spectral content of the marker with the narrower bandwidth is subsumed in the bandwidth of the broader marker (Grose, Hall, Buss, and Hatch, 2001). Furthermore, differences in the number of frequency bands in the leading and lagging markers can be highly disruptive to gap detection (Formby, Sherlock, and Li, 1998). Clearly, these factors are likely to be important when gap cues are used in speech perception.

Other effects of marker asymmetry have been found for differences in intensity level or the fundamental frequency of harmonic complexes (Oxenham, 2000) or for spatial separation (Phillips, Hall, Harrington, and Taylor, 1998; Boehnke and Phillips, 1999). Therefore, various dimensions of dissimilarity between the markers increase the difficulty of detecting a gap between the markers and the effects of these dimensions on gap detection seem to be additive (Taylor, Hall, Boehnke, and Phillips, 1999) and may be related to higher-level processing based on perceptual distance (Chatterjee, Fu, and Shannon, 1998). In "within-channel" conditions (spectrally identical leading and lagging markers), the perceptual operation is thought to involve relatively simple processing of activity in the neural channel representing the stimulus. However, in "between-channel" conditions (spectral differences between the leading and lagging markers), a number of possible factors may be involved: the task is thought to require a relative timing operation between different neural regions (Grose, 1991; Phillips *et al.*, 1997), or to involve a second process reflecting comparisons of components across two or more frequency channels (Viemeister and Plack, 1993), or to be triggered by multiple within-channel decisions (Formby, Sherlock, and Forrest, 1996; Forrest and Formby, 1996; Heinz, Goldstein, and Formby, 1996), or to require cognitive ability (van Wieringen and Wouters, 1999). Speech perception may involve both "within-channel" and "between-channel" processes and age-related differences may involve one and/or the other.

If the specific temporal processing mechanisms involved in detecting a gap when the markers are spectrally asymmetrical are more central than those involved in detecting a gap when the markers are spectrally symmetrical, then it is possible that greater age effects on gap detection thresholds might be observed for the asymmetrical than for the symmetrical marker conditions and that the size of the gap found with spectrally asymmetrical nonspeech markers might be equivalent to the size of the gap necessary for word-pair discrimination. The purpose of the present study is to test this hypothesis by directly comparing the gap detection performance of younger and older listeners using speech and spectrally analogous nonspeech markers varying in both duration and spectral symmetry. The aim of the study was not only to understand better how nonspeech and speech gap detection measures might be related, but also to examine the pattern of age-related differences across the different conditions and how the mechanisms involved in "within-channel" and "between-channel" auditory temporal processing might be differentially affected by aging.

II. METHODS

A. Participants

Sixteen younger adults (aged 21–35 years old; mean =24; SD=3.7) and sixteen older adults (aged 67–82 years old; mean=75; SD=4.4) participated in the study. Participants from the younger group were university students. Participants from the older group were recruited from the local community and all had at least 12 years of formal education. All listeners were native English speakers and in good health with clinically normal pure-tone air-conduction thresholds ($\leqslant$25 dB HL) from 0.25 to 3 kHz in the both ears (see Table I). All participants provided informed consent and were paid an honorarium of $10 per h of testing.

B. Stimuli

There were eight sets of gap detection stimuli. Half of the stimuli were speech and the other half were analogous nonspeech stimuli. The speech and nonspeech stimuli varied in two additional dimensions: marker duration and spectral symmetry. The duration for both the leading and lagging markers was either short (40 ms) or long (250 ms). For spectrally symmetrical stimuli, the leading and lagging markers

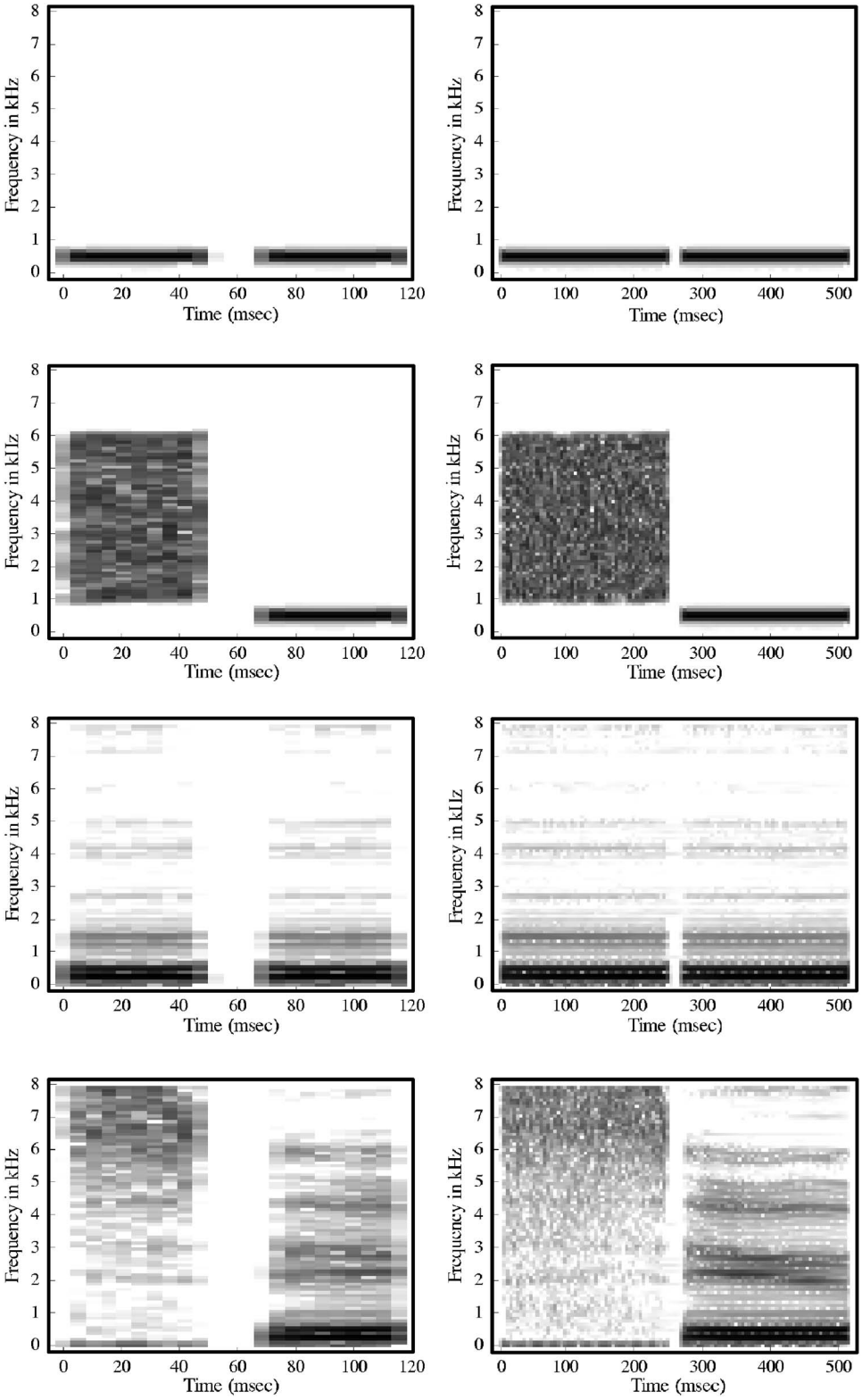


FIG. 1. Spectrograms of the gap stimuli. Short duration stimuli appear in the left column and long duration stimuli in the right column. In each column, the stimuli used in the following conditions are shown from top to bottom: symmetrical nonspeech, asymmetrical nonspeech, symmetrical speech, asymmetrical speech.

were identical in all respects, including spectral content and duration. For the spectrally asymmetrical stimuli, the leading and lagging markers differed in spectral composition but not in duration. For spectrally asymmetrical nonspeech and speech stimuli, the leading marker was a broadband signal, respectively either a 1–6 kHz broadband noise or the consonant [s], and the lagging marker was more tonal, respectively, either a 500-Hz tone or the vowel [u]. The spectral characteristics of the eight stimuli are illustrated in the spectrograms shown in Fig. 1 and their time waveforms are illustrated in Fig. 2.

Nonspeech stimuli were constructed digitally at a sampling rate of 20 kHz and converted to analog using a Tucker Davis Technology (TDT) System III digital-to-analog converter. The 40 ms amplitude envelope was constructed by summing 81 Gaussian envelopes (standard deviation

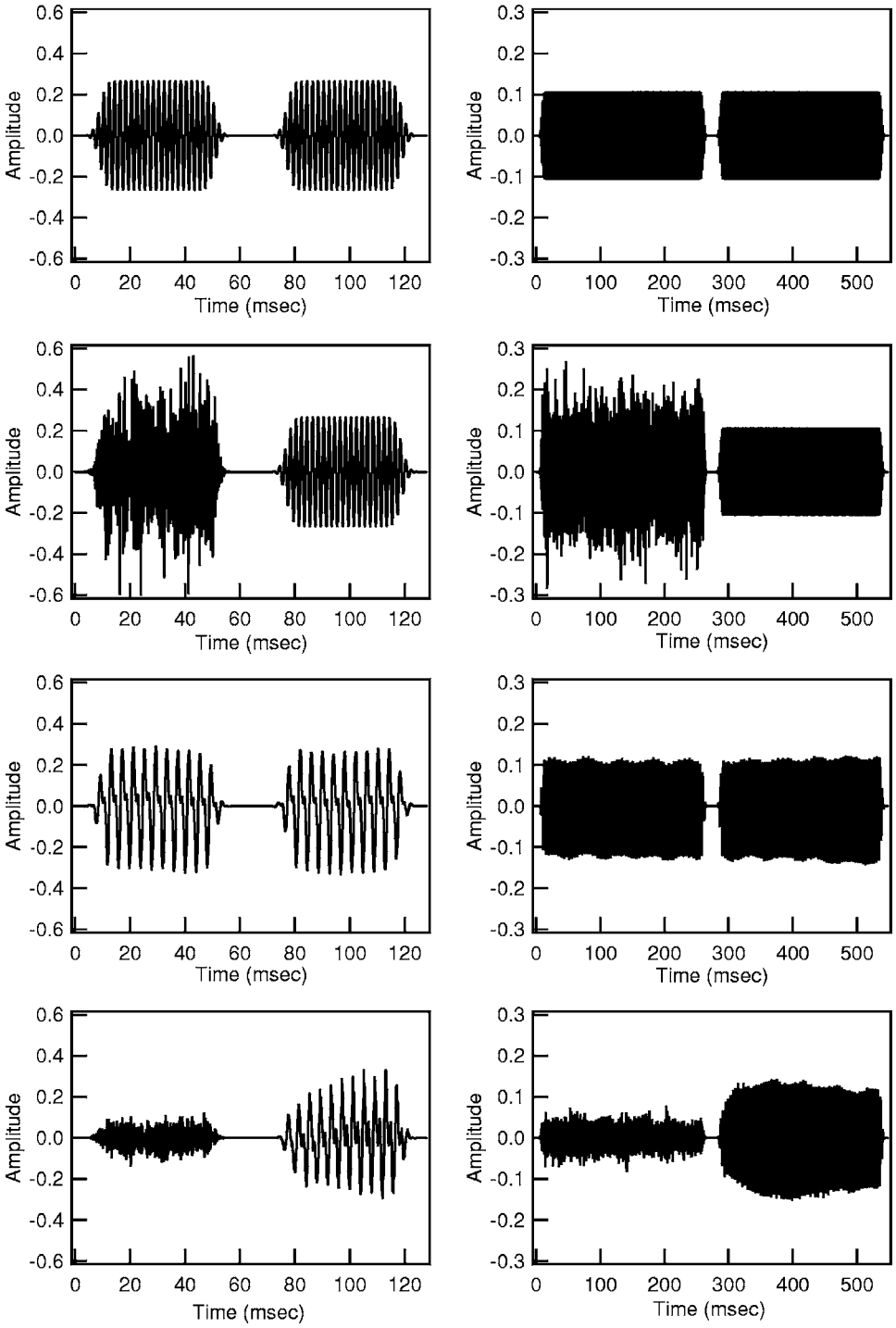


FIG. 2. Time waveforms of the gap stimuli. Short duration stimuli appear in the left column and long duration stimuli in the right column. In each column, the stimuli used in the following conditions are shown from top to bottom: symmetrical nonspeech, asymmetrical nonspeech, symmetrical speech, asymmetrical speech.

=0.5 ms) whose means were spaced 0.5 ms apart, producing an envelope that was flat between 2 and 38 ms, with symmetrical rise and fall times whose shape was determined by the sum of the first five and last five Gaussians in the summation. The 250 ms amplitude envelope was formed by summing 501 Gaussians spaced 0.5 ms apart, producing an envelope that was flat between 2 and 248 ms, with the same rise and fall times as the 40 ms envelope (for a more detailed description of this method see Schneider and Hamstra, 1999). In the nonspeech symmetrical condition, the leading and lagging markers were obtained by multiplying a 500 -Hz tone by either the 40 or 250 ms amplitude envelope. In the nonspeech asymmetrical condition, the lagging marker was produced in the same manner as the markers for the symmetrical condition. The leading marker for the nonspeech asymmetrical condition was produced by multiplying a bandpass white noise (1–6 kHz) by either the 40 or 250 ms amplitude envelope.

The speech stimuli were constructed from recorded samples of [su] and [u] spoken by an adult female. The samples were recorded using a Sennheiser model K3U microphone positioned approximately 6 in. from the talker's mouth and they were digitized at a sampling rate of 20 kHz in a double-walled, sound-attenuating IAC booth using Computer Speech Research Environment 4.5 software (CSRE 4.5, 1995) on a TDT System II. Gaps of varying durations

were inserted in the speech samples at an appropriate zero-crossing in the time waveform between the consonant and vowel of [su] or midway in the steady-state portion of the vowel [u]. The durations of the leading and lagging markers were adjusted by editing at an appropriate zero-crossing at the beginning and end of each utterance. The offsets and onsets of the speech markers were shaped with the same amplitude envelopes (either 40 or 250 ms) used to shape the onsets and offsets of the nonspeech stimuli.

As in our previous studies, the gap in all cases was defined as the time between the end of the leading marker (the location of the peak of the last Gaussian), and the beginning of the lagging marker (location of the peak of the first Gaussian) and the duration of the marker was the time between the peaks of the first and last Gaussian envelopes in a marker (e.g., Schneider and Hamstra, 1999). The Gaussian envelopes were applied to the gap stimuli such that any phase relationships were preserved between the leading and lagging markers (for a discussion of phase effects on gap detection thresholds, see Shailer and Moore, 1987).

For both the symmetrical and asymmetrical conditions, the durations of the control nongap stimuli were matched to the durations of the corresponding test gap stimuli. For the symmetrical stimuli, the gap and control nongap stimuli were matched for overall stimulus duration. Consequently, for the symmetrical stimuli, the duration of a control stimulus was the same as the duration of the two markers plus the duration of the gap in the corresponding gap stimulus. Note that in the symmetrical conditions, the control stimulus was heard as one continuous sound so differences in the duration of the individual markers did not seem to be a potentially useful cue. Because the leading and lagging markers in the asymmetrical stimuli were heard as two distinct sounds even when there was no gap between them, differences in the duration of the two individual markers of the control and gap stimuli might have provided a useful cue. Therefore, a different approach was used to control duration in the asymmetrical conditions (for a discussion of the options for controlling duration see Formby and Muir, 1989). Rather than matching the overall duration of the gap and the nongap control stimuli, the duration of each marker was matched. Consequently, for the asymmetrical stimuli, the duration of the individual markers was eliminated as a possible cue, but total duration was longer for the gap stimuli than for the corresponding control nongap stimuli, with the difference being due to the duration of the gap between the markers. Therefore, in the asymmetrical conditions, the test and control stimuli might have been differentiated either on the basis of the gap or on the basis of a difference in total duration.

For all conditions, the total energy in the control nongap stimuli was matched to the corresponding test gap stimuli. The symmetrical gap detection stimuli were attenuated or amplified in order to match the total energy content of the corresponding markers used in the gap and control conditions. The spectral splatter produced by the introduction of a gap was minimized by the use of the Gaussian envelope to shape onsets and offsets and the availability of spectral splatter as a cue to gap detection was minimized (see the Appendix). All stimuli were presented bilaterally at about 75 dB SPL.

C. Procedure

Gap detection thresholds were measured for all participants in eight conditions (2 marker durations $\times$ 2 symmetry conditions for both speech and nonspeech stimuli). Testing was completed in two sessions of about 1–2 h each, with breaks provided during the sessions. Younger and older participants were randomly assigned to one of eight sequences of presentation of the conditions. Half of the participants were tested first on the short-duration marker conditions and the other half were tested first on the long-duration marker conditions. The order of speech versus nonspeech and symmetrical versus asymmetrical conditions was varied across participants.

A two-interval, two-alternative forced-choice staircase procedure was used. In each trial, listeners heard a gap stimulus in one interval and a nongap stimulus in the other interval, with random assignment of the test gap stimulus and control nongap stimulus to the intervals. Intervals were separated by 1 s. Each interval was marked by the illumination of a light corresponding to one of two buttons on a button box. Participants responded by pressing the button they believed corresponded to the interval in which the gap stimulus was heard. Following each response, feedback was provided by illumination of the light above the correct interval. For the first trial, the size of the gap to be detected depended on the marker symmetry of that condition and was chosen based on pilot experiments. For all symmetrical marker conditions, the initial gap size was 66 ms, and for asymmetrical marker conditions the initial gap size was 350 ms. A 3-down 1-up rule was used to determine the gap size in subsequent trials. For the symmetrical markers, the starting step size was 16 ms and for asymmetrical stimuli it was 32 ms. With each reversal, the step size changed by a factor of 0.5 until a minimum step size of 1 ms was reached. This staircase procedure was used to find the 79.9% point on the psychometric function (Levitt, 1971). Test runs were completed after 12 reversals, and the average of the last eight reversals represented the gap detection threshold for each run. All participants completed at least three runs in each of the eight stimulus conditions. If the thresholds for the participant were continuing to improve between the second and the third run, then up to three more runs were completed until a plateau in performance was observed. The plateau was reached when there was no further improvement in threshold or when the difference between the best two runs was smaller than one standard deviation (based on the distribution of gap thresholds obtained by the younger group in the test condition). About half of the participants in each age group completed extra runs. Gap detection thresholds in each condition were calculated as the mean threshold of the best two runs.

Before each test condition, the participant was first given the instructions and then a practice run using the initial gap size. Further instruction and practice was provided if the lis-

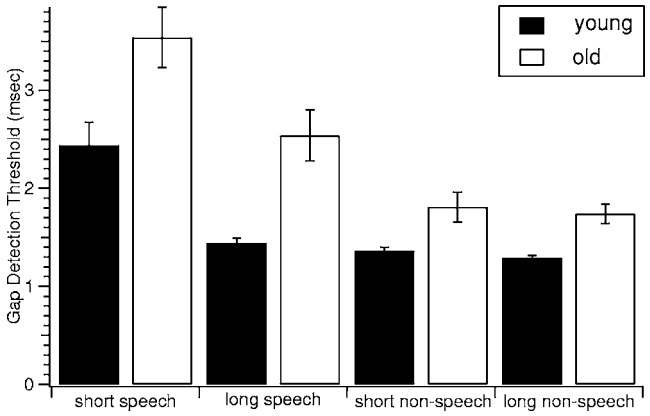


FIG. 3. Mean gap detection thresholds for the younger (black) and older (white) group in short- and long-duration, nonspeech and speech, symmetrical marker conditions. Standard errors are shown.

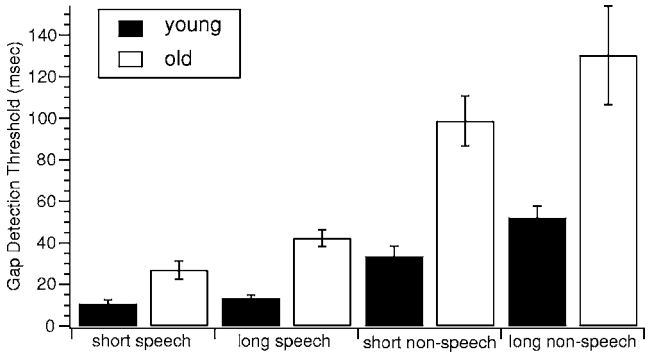


FIG. 4. Mean gap detection thresholds for the younger (black) and older (white) group in short- and long-duration, nonspeech and speech, asymmetrical marker conditions. Standard errors are shown.

tener had difficulty producing six consecutive correct responses on the practice run. The final instructions for the test condition were provided after the practice criteria had been met, followed by an additional practice run using gap sizes that were smaller than in the initial practice runs. Only then were the test runs administered to determine gap detection threshold for the condition.

All testing was performed in a double-walled sound-attenuating IAC booth. Presentation of the stimuli was controlled via MatLab software (version 6.5). Gap detection stimuli were routed from the computer to a TDT System III. The gap detection stimuli were presented binaurally through matched Sennheiser HD-265 headphones via the TDT HB7 headphone driver.

III. RESULTS

Gap detection thresholds for the older group were larger than those of the younger adult group in all eight conditions. For both age groups, gap detection thresholds were much smaller and less variable in spectrally symmetrical conditions than in spectrally asymmetrical conditions (overall mean gap detection threshold in symmetrical conditions =2.0 ms; overall mean gap detection threshold in asymmetrical conditions=51.0 ms). Consequently, separate analyses were conducted for the symmetrical and asymmetrical conditions.

In the spectrally symmetrical conditions shown in Fig. 3, gap detection thresholds were larger for speech than for nonspeech stimuli. For the speech stimuli, gap detection thresholds were larger for short-duration than long-duration markers. An effect of marker duration is not apparent for the nonspeech stimuli. In the short-duration speech condition, gap detection thresholds were larger than in the other symmetrical conditions for both age groups, and the largest age-related difference (1.1 ms) was observed in this condition. In general, age-related differences were greater for speech (1.1 ms) than for nonspeech stimuli (0.4 ms). However, the extent of the age difference did not vary with marker duration. Thus, type of marker contributed to age-related differences but marker duration did not.

This pattern of results was confirmed by a 3 factor ANOVA with age as a between subjects factor, and stimulus type (speech versus nonspeech) and duration as within subject variables. Significant main effects were found for age [$F(1,30)=19.61$, $p<0.001$], stimulus type [$F(1,30)=47.40$, $p<0.001$], and duration [$F(1,30)=37.04$, $p<0.001$]. There were also significant interactions of age x stimulus type [$F(1,30)=5.68$, $p<0.025$], and stimulus type x duration [$F(1,30)=29.37$, $p<0.001$]. No other effects were significant. Student Newman Kuels (SNK) tests of multiple comparisons to examine the age x stimulus type interaction confirmed that the age difference was significant for both speech and nonspeech conditions, with the size of the age difference in gap detection thresholds being larger in the speech conditions ($p=0.01$). than in the nonspeech conditions ($p=0.05$). A SNK test of multiple comparisons to examine the stimulus type x duration interaction confirmed that gap detection thresholds were larger for short-duration than for long-duration speech markers which were in turn significantly larger than gap detection thresholds for nonspeech markers ($p=0.01$), but there was no significant difference between short-duration and long-duration nonspeech markers ($p>0.1$).

The gap detection thresholds obtained in the asymmetrical conditions shown in Fig. 4 are between 4 to more than 50 times larger than those obtained in the corresponding symmetrical conditions shown in Fig. 3. In addition, the pattern of results for the asymmetrical conditions shown in Fig. 4 is reversed compared to the pattern seen for the symmetrical conditions illustrated in Fig. 3. Specifically, mean gap detection thresholds were larger for nonspeech than for speech stimuli (the reverse was true for the symmetrical conditions) and they were larger for long-duration than for short-duration markers (the reverse was true for the symmetrical conditions). For both age groups, the gap detection thresholds for nonspeech stimuli were about 3 times larger than the mean gap detection thresholds for the corresponding speech stimuli. Duration had less effect than the type of stimulus, with the mean gap detection thresholds in the long-duration marker conditions being less than 50% larger than in the corresponding mean gap detection thresholds for short-duration marker conditions. In the long-duration nonspeech condition, gap detection thresholds were the largest for both

age groups and the largest age-related differences in mean gap detection thresholds were observed (78.2 ms). The age-related differences in mean gap detection thresholds were about 50 ms greater for nonspeech than for corresponding speech conditions. Age-related differences in mean gap detection threshold were about 13 ms greater for long-duration markers than for corresponding short-duration markers. Clearly, age-related differences were highly influenced by whether the markers were speech or nonspeech stimuli but duration was of less importance.

An ANOVA confirmed that the same effects that were significant in the symmetrical conditions were also significant in the asymmetrical conditions, except that there was no significant interaction of stimulus type x duration in the asymmetrical conditions. Specifically, there were significant main effects of age [$F(1,30)=27.91$, $p<0.001$], stimulus type [$F(1,30)=49.95$, $p<0.001$], and duration [$F(1,30)=13.04$, $p<0.001$], as well as a significant interaction of age x stimulus type [$F(1,30)=9.86$, $p<0.005$]. The SNK test of multiple comparisons again confirmed that younger adults obtained smaller gap detection thresholds than older adults in both speech and nonspeech conditions, but with the age-related difference being larger in the nonspeech ($p=0.01$) than in the speech conditions ($p=0.05$).

It is important to note that the direction of the main effects of stimulus type and duration were reversed depending on whether the markers were spectrally symmetrical or asymmetrical; whereas gap detection thresholds were smaller for nonspeech than for speech markers in symmetrical conditions, in the asymmetrical conditions gap detection thresholds were smaller for speech than for nonspeech markers; whereas gap detection thresholds were smaller for long-duration than for short-duration markers in symmetrical conditions, in the asymmetrical conditions thresholds were smaller for short-duration than for long-duration markers.

In summary, younger adults detect gaps better than older adults, but both age groups have more difficulty when the gap occurs between spectrally asymmetrical markers than when it occurs between spectrally symmetrical markers. For spectrally symmetrical stimuli, it is *most difficult* to detect a gap between short-duration speech markers and the *greatest* age difference is observed for speech-marker conditions. Conversely, for spectrally asymmetrical stimuli, it is easiest to detect a gap between short-duration speech markers and the *smallest* age difference is observed in this condition. Whether the marker is speech or nonspeech and whether the duration of the marker is 40 ms or 250 ms affects the size of the gap detection thresholds for both age groups; however, the effect due to stimulus type (speech vs nonspeech) is larger than the marker duration effect.

IV. DISCUSSION

In the present study, we set out to compare the gap detection thresholds of younger and older listeners using speech and spectrally analogous nonspeech markers varying in duration and spectral symmetry. The aim of the study was not only to better understand how nonspeech and speech gap detection measures might be related, but also to examine the pattern of age-related differences across the conditions and how the mechanisms involved in "within-channel" and "between-channel" auditory temporal processing might be differentially affected by aging.

The findings confirm that older adults with clinical normal audiograms in the speech range have more difficulty than younger adults detecting gaps in all conditions. The pattern of gap detection thresholds over the set of conditions tested is the same for younger and older adults. For both age groups, gap detection is far more difficult when the leading and lagging markers are spectrally asymmetrical than when they are symmetrical. In the spectrally symmetrical conditions that were tested, the short-duration speech conditions were the most difficult, with the type of stimulus having more influence than the duration of the marker. Curiously, in the asymmetrical conditions, the long-duration nonspeech conditions were by far the most challenging. Even though the nonspeech stimuli used in the present experiments were constructed to be acoustically analogous to the corresponding speech stimuli, gap detection thresholds differed significantly depending on whether the signal was nonspeech or speech. In symmetrical conditions, gaps were detected more easily for nonspeech stimuli; in asymmetrical conditions, gaps were detected more easily for speech stimuli. The findings concerning the effect of the spectral symmetry of the markers, the effect of marker duration, and the effect of stimulus type (speech vs nonspeech) on gap detection thresholds in younger and older listeners are discussed further below.

A. Spectral symmetry

The present findings are in general agreement with previous findings concerning the effect of the spectral symmetry of nonspeech markers on gap detection thresholds. When the spectral composition of the nonspeech markers bounding the gap is identical, younger adult listeners can detect a much smaller gap than when the spectral composition of the markers surrounding the gap are different (e.g., Formby *et al.*, 1996; Phillips *et al.*, 1997; Grose *et al.*, 2001b). In the present study, the average size of the gaps detected by younger listeners when the markers were spectrally symmetrical was 1.3 ms for nonspeech markers; however, they were much larger in the asymmetrical conditions, ranging from 33 ms for the short-duration nonspeech markers up to 52 ms for the long-duration nonspeech markers. The present findings are similar to the typical finding of gap detection thresholds around 5 ms in other studies using similar spectrally symmetrical nonspeech markers (e.g., Shailer and Moore, 1987; Moore and Glasberg, 1988; Formby and Forrest, 1991). The present findings are also similar to typical findings of larger gap thresholds of 10 ms to over 50 ms, depending on the degree of spectral asymmetry, in other studies using similar asymmetrical nonspeech markers, (e.g., Phillips *et al.*, 1997; Formby *et al.*, 1996).

The effect of the spectral symmetry of the markers on the gap detection thresholds of older adults was similar to that of younger adults; however, the gap detection thresholds of the older adults in the present study were about 30%

larger than those of the younger adults in the nonspeech spectrally symmetrical marker conditions and up to about 200% larger in the spectrally asymmetrical nonspeech marker conditions. The present results for both age groups in the symmetrical nonspeech condition agree with earlier studies conducted by our group (e.g., Schneider *et al.*, 1994; Schneider *et al.*, 1998; Schneider and Hamstra, 1999). The present findings for older listeners in the asymmetrical conditions are consistent with previous studies in which age-related differences were found to be much larger for spectrally asymmetrical than for spectrally symmetrical 250-ms $\frac{1}{4}$-octave band noise markers (Lister *et al.*, 2002) and with another study study in which age-related differences were found for gaps between synthetic speech markers, with gap discrimination being more difficult when the markers were spectrally dynamic compared to when they were more stable (Lister and Tarver, 2004). Other studies comparing gap thresholds for normal-hearing compared to hearing-impaired adults have also suggested that age-related declines seem to be more pronounced when nonspeech markers are spectrally asymmetrical (Grose *et al.*, 2001b; Lister *et al.*, 2002). Consistent with previous studies of the effect of spectral symmetry and gap discrimination, the present study provides new evidence that age-related differences in ability to detect a gap in spectrally simple and symmetrical markers are exacerbated when the markers bounding the gap are more spectrally complex and asymmetrical, as they are in speech.

B. Marker duration

The differences in marker duration tested in the present study (40-ms vs 250-ms markers) had much less effect on gap detection thresholds than did spectral symmetry. Previous studies of the effect of marker duration have shown that "between channel" gap detection thresholds are larger when the duration of the leading marker is less than about 30 ms and that performance asymptotes when the marker is longer (e.g., Phillips *et al.*, 1997). Previous studies have also found that age-related differences in gap detection thresholds are significant when marker duration is less than 40 ms but not when the marker duration is longer (Schneider and Hamstra, 1999). The short marker duration of 40-ms was chosen for the present study to provide a difficult condition in which age and "between-channel" effects might be observed, whereas the long marker duration of 250-ms was chosen to provide an easy condition in which asymptotic performance could be measured with minimal age-differences. Although the 40-ms marker duration was near the limit of the range of leading marker durations where age and "between channel" effects have been observed, because our intention was to test marker durations that might occur in speech, 40 ms was the shortest marker duration that could be used without disrupting the perception of the stimuli as speech. Unfortunately, the weak effects of marker duration on gap detection thresholds may have resulted because performance in the 40-ms marker duration conditions was near asymptote. Nevertheless, consistent with previous findings, gap detection thresholds were slightly smaller for long-duration than for short-duration spectrally symmetrical markers (e.g., Schneider and Hamstra, 1999). Age-related differences were equivalent in the short-duration and long-duration versions of the symmetrical and asymmetrical marker conditions. The weak effects of marker duration on gap detection threshold were also similar for both speech and nonspeech markers. A surprising finding was that gap detection thresholds were markedly larger for long-duration than for short-duration spectrally asymmetrical markers; however, overall, the effects of marker duration on gap detection thresholds are of only minor interest compared to the other results of the present study.

C. Stimulus type (speech vs nonspeech)

Although previous studies have investigated the correlation between gap detection thresholds for nonspeech markers and speech perception (e.g., Tyler *et al.*, 1982; Snell *et al.*, 2002) and some investigators have chosen nonspeech gap markers that approximated the spectral properties of consonant-vowel utterances (e.g., Formby *et al.*, 1993; Phillips *et al.*, 1997), to our knowledge gap detection thresholds in actual speech markers have not been compared to gap detection thresholds for analogous nonspeech markers. Even though the patterns of results are similar for speech and analogous nonspeech markers, some interesting differences related to stimulus type were observed. When the markers surrounding the gap were spectrally identical, performance was poorer for speech than for nonspeech markers. The speech marker was a naturally spoken vowel [u] and the nonspeech marker was a 500-Hz Gaussian-enveloped tone-pip. Both marker types were periodic and spectrally steady; however, the vowel marker was more spectrally complex and somewhat less periodic than its tonal counterpart. The greater spectral complexity of the vowel markers seems unlikely to be the explanation considering that gap detection thresholds have been found to improve when the number of frequency bands in symmetrical markers is increased (e.g., Grose *et al.*, 1996; Phillips and Hall, 2000). It seems more likely that the degree of periodicity of the marker may have influenced the detectability of the gap because gap detection thresholds are known to be vulnerable to phase disruptions (e.g., Shailer and Moore, 1987).

When the markers surrounding the gap were spectrally asymmetrical, performance was poorer for nonspeech than for speech markers. Four possible factors may explain the differences: spectral overlap of the markers, the distribution of energy across frequencies, the use of other speech-specific cues, and/or phonological knowledge.

The first possibility is that the degree of spectral overlap of the markers enhanced the detection of gaps in asymmetrical gap stimuli. In the nonspeech case, the leading marker was a broadband noise (1–6 kHz) and the lagging marker was a 500-Hz tone. In the speech case, the leading marker was [s] and the lagging marker was [u]. There was no spectral overlap between the nonspeech markers; however, there was some spectral overlap between the speech markers. While this small degree of spectral overlap may have been beneficial, the extent of such benefit is open to question given that previous studies have shown that spectral dissimi-

larity between markers overrides the possible usefulness of available within-channel information (Phillips and Hall, 2002).

The second possibility is that the distribution of energy across frequencies enhanced performance in asymmetrical speech conditions compared to nonspeech conditions. In the nonspeech case, the energy in the leading broadband noise marker was evenly distributed over the 1–6 kHz frequency range. In the speech case, the energy of the [s] spanned a broader range of frequencies (up to 8 kHz), but the energy was more concentrated in the high frequencies than the low frequencies. The presence of higher frequency components in low-pass noise markers results in a reduction in gap detection thresholds for both young and old listeners (e.g., Snell, 1997). If the ability to detect a gap were enhanced by a broader range of frequency components or by a concentration of energy in the higher frequencies then gap thresholds might be expected to be smaller for the asymmetrical speech than for the asymmetrical nonspeech conditions, with the younger adults benefiting more than older adults who had elevated audiometric thresholds at higher frequencies. The finding that age-related differences were less rather than more pronounced in the asymmetrical speech compared to the asymmetrical nonspeech conditions is not consistent with the distribution of energy across frequencies as an explanation.

A third possible explanation is that other useful acoustic cues were available in the speech stimuli that were not available in the nonspeech stimuli. Although we focused on the spectral properties of the nonspeech and speech markers, because the speech markers were constructed from naturally spoken speech, there may have been other helpful acoustical cues, such as the periodicity cues arising from the fundamental frequency and harmonic structure of the voice. Participants reported hearing the nonspeech stimuli as a set of two distinct sounds but listeners were sometimes confused about the order of the sounds. The speech markers were clearly heard as speech and the order of the sounds was not confused. The acoustic cues related to voice perception may have conferred an advantage on gap detection. This possibility may be consistent with the benefit to gap detection of continuity between the leading and lagging markers in terms of preserved phase (Shailer and Moore, 1987) and constant F0 (Oxenham, 2000). Note that although phase was preserved in the periodic nonspeech and speech markers used in the symmetrical conditions, in the asymmetrical conditions the noise band leading markers were aperiodic, but the [s] leading markers may have retained a small degree of advantageous voice periodicity.

A final possibility is that performance was better in the spectrally asymmetrical speech conditions than in the analogous nonspeech conditions because listeners were able to deploy phonological knowledge to advantage when processing the signals. Hearing an unfamiliar sequence of unrelated unnatural sounds such as a noise followed by a tone may invoke very different higher-level auditory processing than hearing a sequence resembling a highly familiar string of speech sounds spoken by a single voice. This possibility is consistent with the suggestion that linguistic experience may alter the perceptual (re-)organization of speech stimuli by listeners with different degrees of cochlear hearing loss (Nelson *et al.*, 1995).

The size of the age-related difference in gap detection thresholds in spectrally asymmetrical nonspeech markers was over 70 ms but it was less than 25 ms for asymmetrical speech markers. Even though it was more difficult for both age groups to detect gaps between spectrally asymmetrical markers than gaps between spectrally symmetrical markers, and even though the age-related differences were much larger in the more difficult spectrally asymmetrical marker conditions, it is very interesting that detecting a gap between speech markers was so much less difficult than detecting a gap between nonspeech markers. It is even more interesting that older adults were relatively immune to the difficulty of detecting a gap between spectrally asymmetrical markers when the markers were speech rather than nonspeech stimuli.

The reduced ability of older listeners with good audiograms in the speech range to detect gaps between simple, spectrally identical markers is now well established, with a possible explanation being age-related differences in recovery from adaptation (Schneider and Hamstra, 1999). Note that, even though the older participants had elevated audiometric thresholds at high frequencies (above 3 kHz), their difficulty in detecting gaps could not be predicted from their degree of high-frequency hearing loss (4, 6, or 8 kHz thresholds). Specifically, out of a total of 48 correlations (three frequencies $\times$ two ears $\times$ 8 gap thresholds), there were only four correlations that reached significance, of which two were positive but two were negative. Therefore, consistent with our previous studies (e.g., Schneider *et al.*, 1994), there is no convincing evidence that the gap detection thresholds of the older listeners in the present study are attributable to their high-frequency pure-tone audiometric thresholds.

In addition to age-related differences in the processing of discontinuities "within channel," the present findings suggest that there are also age-related differences in the more "central" or "between channel" processes that are required to detect gaps between more complex and spectrally asymmetrical markers. The specific nature of the mechanisms involved when a gap is detected between more complex markers is not yet understood; however, many investigators have assumed that in addition to the processing involved in "within channel" gap detection, there must be a second process that differs perceptually and physiologically when a gap is detected between events processed in different "channels" (Viemeister and Plack, 1993). A multichannel model involving peripheral filtering and a within-channel decision metric has been proposed and tested in computer simulations by Formby and co-workers (Formby *et al.*, 1996; Heinz *et al.*, 1996). Phillips and co-workers have hypothesized that a "central" relative timing operation is involved when different channels are activated before and after a gap (e.g., Phillips *et al.*, 1997; Taylor *et al.*, 1999; Phillips and Hall, 2000). Higher-level perceptual or cognitive hypotheses involving ability to attend to the gap and ignore distracting differences in complex stimuli have also been put forward (e.g., Formby *et al.*, 1998; Chatterjee *et al.*, 1998; van Wieringen and Wouters, 1999). The suggestion of higher-level processing of

gaps is compatible with physiological studies of the brainstem (e.g., Walton, Frisina, and O'Neill, 1998; Werner, Folsom, Mand, and Syapin, 2001) and studies of the auditory cortex that indicate that there is a left hemisphere advantage for the discrimination of temporal events (e.g., Brown and Nicholls, 1997; Sulakhe, Elias, and Lejbak, 2003), and the clear evidence that minimum gaps are represented in the firing pattern of neurons in auditory cortex (Eggermont, 2000).

It may be that peripheral coding of gap information is sufficient for the detection of gaps between simple, spectrally symmetrical markers and that it is a prerequisite to additional processing involved in the detection of gaps between complex, spectrally asymmetrical markers. If minimum gaps are maintained in representations up to the auditory cortex, then it is possible that perceptual mechanisms important in streaming, grouping, and organizing auditory images and patterns may modulate attention to gaps as has been suggested by others (Formby *et al.*, 1998; Phillips and Hall, 2000). To the extent that phonological knowledge facilitates the parsing and organization of the ongoing speech signal, it may confer an advantage for detecting gaps in speech stimuli that does not extend to the auditory processing of nonspeech sounds. Even though older adults have declines in auditory temporal processing that reduce their ability to detect gaps in both "within" and "between" channel marker conditions, they may nevertheless benefit from well-preserved highly overlearned phonological knowledge. Even though auditory processing declines with age, phonological knowledge may be preserved. Preserved phonological knowledge may compensate to some extent for age-related auditory declines. Accordingly, age-related deficits in speech understanding in noise may involve somewhat of a trade off of reduced auditory processing abilities against preserved expert linguistic knowledge, including phonological knowledge. Such a compensatory trade-off between loss of auditory processing and persevered phonological knowledge is reminiscent of the well-known cognitive compensatory trade-off of declines in age-related information processing vs preservation of semantic knowledge.

ACKNOWLEDGMENTS

This research was funded by the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes of Health Research. The first half of the participants were tested in the Master's thesis research of Edward Storzer at UBC. We would like to thank Marko Nabokov, Lesley Filmer, Aaron Jacobs, and Christine DeLuca for assistance with testing the second half of the participants at UTM. We also wish to thank Dariush Ebrahimi for assistance in constructing the stimuli.

APPENDIX

If a gap is created in the middle of a pure tone, there will be some spectral splatter when the gap occurs. There are several reasons to expect that the contribution of spectral splatter to the detection of a gap was minimal in the present study. First, neither the authors nor participants heard clicks when gaps were produced in any of the stimuli. Second, the spectrograms for the stimuli do not show any indication of bursts of energy in the high-frequency region when a gap is introduced. Third, to obtain a more precise estimate of the amount of spectral splatter produced by a gap, we first multiplied the 500-Hz pure tone containing the gap by a 15.5 ms Gaussian window, centered on the gap. We then determined the spectral energy density function of the resulting product. The 15.5 ms Gaussian window consisted of the sum of 32 Gaussians spaced 0.5 ms apart, with each Gaussian having a standard deviation of 0.5 ms. This window is shown in Fig. 5.

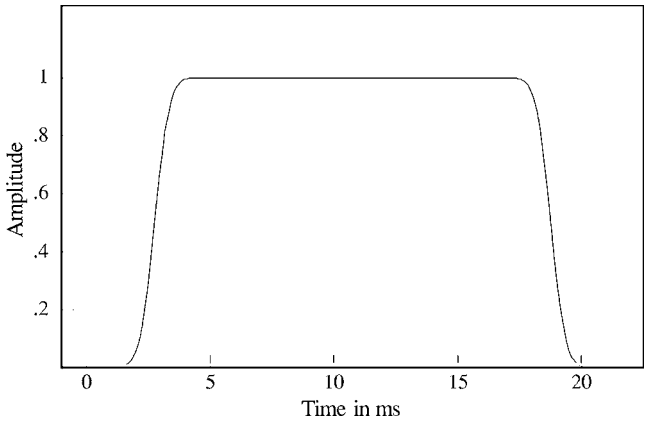


FIG. 5. The Gaussian window used in the spectral analysis in Fig. 6. This window was produced by summing 32 Gaussians spaced 0.5 ms apart with each Gaussian having a standard deviation of 0.5 ms. The duration of the window (15.5 ms) is defined by the time difference between the midpoints of the first and last Gaussians in the sum.

The Gaussian window shown in Fig. 5 was applied to a 500-Hz pure tone without a gap, a 500-Hz pure tone with a 3.5 ms gap, and a 500-Hz tone with a 6.5 ms gap. In the latter two cases, the Gaussian window was centered on the gap. The spectral energy density function for each of these three windowed stimuli is shown in Fig. 6. Fig. 6 illustrates that if the listener is processing the signals using a 15.5 ms window, it is unlikely that she or he will be able to discriminate among these stimuli based on their spectral profiles.

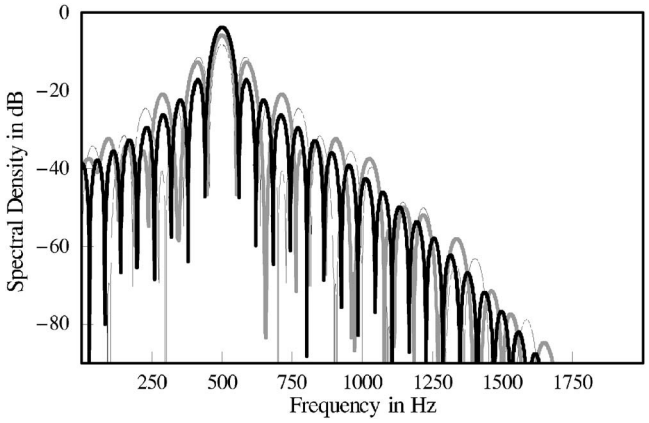


FIG. 6. Energy density in dB as a function of frequency for three windowed stimuli: a 500 Hz pure tone (thick dark line); a 500 Hz pure tone with a 3.5 ms gap (thick grey line); and a 500 Hz pure tone with a 6.5 ms gap (thin dark line). For the first stimulus the Gaussian window (see Fig. 5) was centered in the middle of the pure tone. For the latter two stimuli, the Gaussian window was centered at the midpoint of the gap.

Moreover, Fig. 6 shows that the spectral splatter that is introduced when gaps in pure tones are formed using this method, does not extend up to the high frequency region. Hence, the high-frequency hearing loss of the older adults is unlikely to place them at a disproportionate disadvantage compared to younger adults with respect to the usefulness of any spectral splatter cues in the gap detection task.

Bergman, M. (**1980**). *Aging and Speech Perception*, Perspectives in Audiology Series (University Park Press, Baltimore, MD).

Boehnke, S. E., and Phillips, D. P. (**1999**). "Azimuthal tuning of human perceptual channels for sound location," J. Acoust. Soc. Am. **106**, 1948–1955.

Brown, S., and Nicholls, M. (**1997**). "Hemispheric asymmetries for the temporal resolution of brief auditory stimuli," Percept. Psychophys. **59**, 442–447.

CHABA (Committee on Hearing, Bioacoustics, and Biomechanics) (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859–895.

Chatterjee, M., Fu, Q.-J., Shannon, R. V. (**1998**). "Within-channel gap detection using dissimilar markers in cochlear implant listeners," J. Acoust. Soc. Am. **103**, 2515–2519.

Divenyi, P., and Simon, H. (**1999**). "Hearing in aging: Issues old and young," Curr. Opin. Otolaryngol.-Head Neck Surg. **7**, 282–289.

Eddins, D., Hall, J. W., and Grose, J. (**1992**). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," J. Acoust. Soc. Am. **91**, 1069–1077.

Eggermont, J. J. (**2000**). "Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap detection thresholds," J. Neurophysiol. **84**, 1453–1463.

Florentine, M., and Buus, S. (**1984**). "Temporal gap detection in sensorineural and simulated hearing impairments," J. Speech Hear. Res. **27**, 449–455.

Formby, C., Barker, C., Abbey, H., and Raney, J. J. (**1993**). "Detection of silent temporal gaps between narrow-band noise markers having second-formant-like properties of voiceless stop/vowel combinations," J. Acoust. Soc. Am. **93**, 1023–1027.

Formby, C., and Forrest, T. G. (**1991**). "Detection of silent temporal gaps in sinusoidal markers," J. Acoust. Soc. Am. **89**, 830–837.

Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (**1998**). "Evidence for an across-frequency, between-channel process in asymptotic monaural gap detection," J. Acoust. Soc. Am. **103**, 3554–3560.

Formby, C., and Muir, K. (**1989**). "Effects of randomizing signal level and duration on temporal gap detection," Audiology **28**, 250–257.

Formby, C., Sherlock, L. P., and Forrest, T. G. (**1996**). "An asymmetric roex filter model for describing detection of silent temporal gaps in sinusoidal markers," Aud. Neurosci. **3**, 1–20.

Formby, C., Sherlock, L. P., and Li, S. (**1998**). "Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position," J. Acoust. Soc. Am. **104**, 984–998.

Forrest, T. G., and Formby, C. (**1996**). "Detection of silent temporal gaps in sinusoidal markers simulated with a single-channel envelope detector model," Aud. Neurosci. **3**, 21–33.

Frisina, D. R., and Frisina, R. D. (**1997**). "Speech recognition in noise and presbycusis: Relations to possible neural sites," Hear. Res. **106**, 95–104.

Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808–1817.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.

Grose, J. H. (**1991**). "Gap detection in multiple narrow bands of noise as a function of spectral configuration," J. Acoust. Soc. Am. **90**, 3061–3068.

Grose, J. H., Eddins, D. A., Hall, J. W. III. (**1989**). "Gap detection as a function of stimulus bandwidth with fixed high-frequency cutoff in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **86**, 1747–1755.

Grose, J. H., and Hall, J. W. III. (**1993**). "Gap detection in a narrow band of noise in the presence of a flanking band of noise," J. Acoust. Soc. Am. **93**, 1645–1648.

Grose, J. H., Hall, J. W. III, and Buss, E. (**2001**). "Gap duration in listeners with cochlear hearing loss: Effects of gap and marker duration, frequency separation, and mode of presentation," J. Assoc. Res. Otolaryngol. **2**, 388–398.

Grose, J. H., Hall, J. W. III, Buss, E., and Hatch, D. (**2001b**). "Gap detection for similar and dissimilar gap markers," J. Acoust. Soc. Am. **109**, 1587–1595.

Hall, J. W. III, Grose, J. H., and Joy, S. (**1996**). "Gap detection for pairs of noise bands: Effects of stimulus level and frequency separation," J. Acoust. Soc. Am. **99**, 1091–1095.

Haubert, N., and Pichora-Fuller, M. K. (**1999**). "The perception of spoken language by elderly listeners: Contributions of auditory temporal processes," Can. Acoust. **27**(3), 96–97.

He, N.-J., Horwitz, R., Dubno, J., and Mills, J. H. (**1999**). "Psychometric functions for gap detection in noise measured from young and aged subjects," J. Acoust. Soc. Am. **106**, 966–978.

Heinz, M. G., Goldstein, M. H. Jr., and Formby, C., (**1996**). "Temporal gap detection thresholds in sinusoidal markers simulated with a multichannel, multiresolution model of the auditory periphery," Aud. Neurosci. **3**, 35–56.

Levitt, H. (**1971**). "Transformed up-down methods in psychacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lister, J., Besing, J., and Koehnke, J. (**2002**). "Effects of age and frequency disparity on gap discrimination," J. Acoust. Soc. Am. **111**, 2793–2800.

Lister, J., Koehnke, J., and Besing, J. (**2000**). "Binaural gap duration discrimination in listeners with impaired hearing and normal hearing," Ear Hear. **21**, 141–150.

Lister, J., and Tarver, K. (**2004**). "Effect of age on silent gap discrimination in synthetic speech stimuli," J. Speech Lang. Hear. Res. **47**, 257–268.

Moore, B. C. J., and Glasberg, B. R. (**1988**). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," J. Acoust. Soc. Am. **83**, 1093–1101.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1993**). "Detection of temporal gaps in sinusoids: Effects of frequency and level," J. Acoust. Soc. Am. **93**, 1563–1570.

Nelson, P. B., Nittrouer, S., and Norton, S. J. (**1995**). ""Say-stay" identification and psychoacoustic performance of hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 1830–1838.

Ohde, R. N., and Abou-Khalil, R. (**2001**). "Age differences for stop-consonant and vowel perception in adults," J. Acoust. Soc. Am. **110**, 2156–2166.

Oxenham, A. J. (**2000**). "Influence of spatial and temporal coding on auditory gap detection," J. Acoust. Soc. Am. **107**, 2215–2223.

Phillips, D. P., and Hall, S. E. (**2000**). "Independence of frequency channels in auditory temporal gap detection," J. Acoust. Soc. Am. **108**, 2957–2963.

Phillips, D. P., and Hall, S. E. (**2002**). "Auditory temporal gap detection for noise markers with partially overlapping and nonoverlapping spectra," Hear. Res. **174**, 133–141.

Phillips, D. P., Hall, S. E., Harrington, I. A., and Taylor, T. L. (**1998**). "'Central' auditory gap detection: A spatial case," J. Acoust. Soc. Am. **103**, 2064–2068.

Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," J. Acoust. Soc. Am. **101**, 3694–3705.

Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen to and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.

Pichora-Fuller, M. K., and Souza, P. (**2003**). "Effects of aging on auditory processing of speech," International Journal of Audiology **42** (Supplment 2), S11–S16.

Price, P. J., and Simon, H. J. (**1984**). "Perception of temporal differences in speech by 'normal-hearing' adults: Effects of age and intensity," J. Acoust. Soc. Am. **76**, 405–410.

Sulakhe, N., Elias, L., and Lejbak, L. (**2003**). "Hemispheric asymmetries for gap detection depend on noise type," Brain Cogn **53**, 372–375.

Schneider, B. A., and Hamstra, S. J. (**1999**). "Gap detection thresholds as a function of tonal duration for younger and older listeners," J. Acoust. Soc. Am. **106**, 371–380.

Schneider, B. A., and Pichora-Fuller, M. K. (**2000**). "Implications of perceptual processing for cognitive aging research," in *The Handbook of Aging and Cognition*, 2nd ed., edited by F. I. M. Craik and T. A. Salthouse (Lawrence Erlbaum Associates, New York).

Schneider, B. A., and Pichora-Fuller, M. K. (**2001**). "Age-related changes in temporal processing: Implications for listening comprehension," Semin. Hear. **22**, 227–239.

Schneider, B. A., Speranza, F., and Pichora-Fuller, M. K. (**1998**). "Age-

related changes in temporal resolution: Envelope and intensity effects," Can. J. Exp. Psychol. **52**, 184–190.
Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980–991.
Shailer, M. J., and Moore, B. C. J. (**1987**). "Gap detection and the auditory filter: Phase effects using sinusoidal stimuli," J. Acoust. Soc. Am. **81**, 1110–1117.
Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.
Snell, K. B., and Frisina, D. R. (**2000**). "Relationships among age-related differences in gap detection and word recognition," J. Acoust. Soc. Am. **107**, 1615–1626.
Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). "Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity," J. Acoust. Soc. Am. **112**, 720–727.
Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.
Taylor, T. L., Hall, S. E., Boehnke, S. E., and Phillips, D. P. (**1999**). "Additivity of perceptual channel-crossing effects in auditory gap detection," J. Acoust. Soc. Am. **105**, 563–566.
Tyler, R. S., Summerfield, Q., Wood, E. J., and Fernandes, M. A. (**1982**). "Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **72**, 740–752.
van Wieringen, A., and Wouters, J. (**1999**). "Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees," J. Acoust. Soc. Am. **106**, 1925–1939.
Viemeister, N. F., and Plack, C. J. (**1993**). "Time analysis," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).
Walton, J. P., Frisina, R. D., and O'Neill, W. E. (**1998**). "Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse," J. Neurosci. **18**, 2764–2776.
Werner, L. A., Folsom, R. C., Mand, L. R., and Syapin, C. L. (**2001**). "Human auditory brainstem response to temporal gaps in noise," J. Speech Lang. Hear. Res. **44**, 737–750.
Wingfield, A. (**1996**). "Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory," J. Am. Acad. Audiol **7**, 175–182.

Frequency transposition around dead regions simulated with a noiseband vocoder

Deniz Başkent[a] and Robert V. Shannon
Department of Biomedical Engineering, University of Southern California, Los Angeles, California 90089 and Department of Auditory Implants, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 11 July 2005; revised 22 November 2005; accepted 23 November 2005)

In sensorineural hearing loss, damage to inner hair cells or the auditory nerve may result in dead regions in the cochlea, where the information transmission is disrupted. In cochlear implants, similar dead regions might appear if the spiral ganglia do not function. Shannon *et al.* [J. Assoc. Res. Otolaryngol. **3**, 185–199 (2002)] simulated dead regions of varying size and location using a noiseband vocoder. Phoneme recognition by normal-hearing subjects was measured under two frequency-place mapping conditions: the frequency range corresponding to the dead region was (1) removed or (2) reassigned to bands adjacent to the dead region to simulate the off-frequency stimulation of neurons at the edge of a dead region. The present study extends the results of Shannon *et al.* by including a frequency transposition mapping condition, where the overall acoustic input frequency range was distributed over the entire remaining nondead region. The frequency transposed map provided more acoustic information when compared to the map with the frequency range corresponding to the dead region removed. However, speech perception did not improve for many simulated dead region conditions, possibly due to the spectral distortions in the frequency-place mapping. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151825]

PACS number(s): 43.71.Ky, 43.66.Ts [DOS] Pages: 1156–1163

I. INTRODUCTION

Sensorineural hearing loss is often caused by damage in hair cells or the auditory nerve. When the damage is in the outer hair cells, the active amplification mechanism in the cochlea diminishes and the auditory filters become broader (Glasberg and Moore, 1986). Such reduced frequency selectivity is believed to be one of the main factors leading to poor speech recognition in noise by hearing-impaired listeners (Stelmachowicz *et al.*, 1985). When the damage is in the inner hair cells or in the auditory nerve, the efficiency of the information transmission to higher stages of the auditory pathway is reduced (Moore and Glasberg, 1997). Listeners with this type of pathology usually demonstrate poor speech recognition even in quiet listening conditions (Pauler *et al.*, 1986). In cochlear implants, the hair cells are bypassed and the sound information is delivered by direct electrical stimulation of the auditory nerve. However, the survival rate or the functionality of the spiral ganglia, or abnormalities in more central auditory nerves, might affect speech recognition abilities of cochlear implant users (Gomaa *et al.*, 2003). Histopathological studies on temporal bones taken post mortem from hearing-impaired (Schuknecht and Gacek, 1993) or implant subjects (Khan *et al.*, 2005) showed that the cochlear pathology may differ for each patient. In hearing impairment, if the inner hair cells or the auditory nerve are severely damaged, the transduction from mechanical vibrations of basilar membrane to action potentials might be disrupted for that particular region, resulting in a dead region (Moore and Glasberg, 1997). In implants, if the spiral ganglia did not survive, the transmission of the electrical signal to the nerve might similarly be hindered.

Shannon *et al.* (2002) simulated dead regions with cochlear implant listeners, by selectively turning off electrodes, and in normal-hearing listeners, by using a noiseband vocoder. The analysis bands of the vocoder represented the acoustic input, and the carrier bands represented the cochlear stimulation range. The dead regions were simulated by dropping carrier bands of the vocoder. In the first frequency-place mapping condition, the acoustic information falling in the dead region was also dropped, preserving the matched spectral mapping for the remaining regions. Speech recognition was measured as a function of the size and location of the dead region. The simulated cochlear locations of the dead regions were apical, middle, and basal, with the logarithmic center frequencies of around 1, 2.4, and 5 kHz, respectively. Once normalized for the baseline condition where there was no dead region, performances by implant users and normal-hearing subjects were similar. Speech recognition by all subjects decreased as the size of the dead region increased. The effect was more pronounced for dead regions simulated in the apical region, where frequencies around 1 kHz were missing.

Because there is no or minimal information transmission in a dead region, the audiometric threshold of a tone falling in the dead region would theoretically be infinite. At loud levels, however, the neighboring healthy regions can be stimulated by the spread of basilar membrane vibrations, resulting in an off-place detection of the tone (Florentine and Houtsma, 1983). In this case, even though the correct place for the tone frequency was not stimulated, the audiogram

[a]Electronic mail: deniz_baskent@starkey.com

 0001-4966/2006/119(2)/1156/8/$22.50

would show a finite but high threshold for the tone frequency. Numerous studies (Rankovic, 1991; Hogan and Turner, 1998; Vickers *et al.*, 2001) have indicated that amplification in dead regions might be more harmful than beneficial, probably due to such distortions in the frequency-place excitation patterns. Similar spectral distortions may arise in electrical hearing if an electrode is located in a region where the spiral ganglia do not function or are mostly vanished. The current levels, then, might have to be set to high values to stimulate a sufficient number of spiral ganglia to produce a sensation. Irregularities in the stimulation levels, measured for electrodes located at different cochlear locations, have been shown for stimulation modes that produce a narrow spread of activation, such as bipolar (Pfingst and Xu, 2004; Pfingst *et al.*, 2004) or tripolar (Bierer, 2005) configurations. The strong current fields produced due to high stimulation levels might lead to off-place stimulation of the ganglia on the edges of the dead region. In the second condition, Shannon *et al.* (2002) simulated such off-place stimulation as a local distortion in the frequency-place mapping. They reassigned the envelopes from the analysis bands in the dead region to the carrier bands immediately neighboring the dead region. The results with the reassigned condition did not differ from the results of the first condition, where the bands were simply dropped.

As a feasible alternative to amplification in the dead regions, where there was no benefit for some hearing-impaired subjects, frequency transposition was suggested. The idea was to shift the frequency information from the dead region into an area of residual hearing. Most studies showed no significant improvement in speech recognition with frequency transposition (e.g., McDermott and Dean, 2000). However, Turner and Hurtig (1999) did show a benefit for some hearing-impaired subjects, when they kept the relative distance of formant peaks proportionate in the transposition of the high frequencies.

The dead regions with hearing-impaired listeners limit the stimulation range of the cochlea that can be used for transmission of the acoustic information. A similar problem occurs with cochlear implant users when the stimulation range of the electrode array is limited. For example, if the array is inserted shallower than a full insertion, the apical cochlear regions, which would normally respond to low frequencies, will not be activated by electrical stimulation. Başkent and Shannon (2005) simulated shallow insertions with Med-El 40+ implant users and applied frequency transposition by compressing the entire speech spectrum to the limited stimulation range. An improvement in phoneme recognition was observed when the input frequency range was moderately compressed, but the performance dropped as the compression rate increased.

The present study is an extension of the Shannon *et al.* (2002) study and explores the feasibility of frequency transposition as a method to increase speech information transmission in the presence of dead regions, simulated with a noiseband vocoder. The first mapping condition in Shannon *et al.* (2002), where the acoustic information in the dead region was discarded, was repeated as the baseline condition in the present study. Dead regions of varying size and cochlear location were simulated with a noiseband vocoder and phoneme recognition by normal-hearing listeners was measured. In a second mapping condition, frequency transposition, the entire input spectrum was distributed over the remaining carrier bands that represented the nondead cochlear region. This second map is different from the locally distorted mapping condition simulated by Shannon *et al.*, where the analysis band envelopes from the dead region were assigned only to the carrier bands at the edges of the simulated dead region. In the frequency transposition map used in the present study, the distortion is distributed evenly across the entire nondead region (Fig. 1).

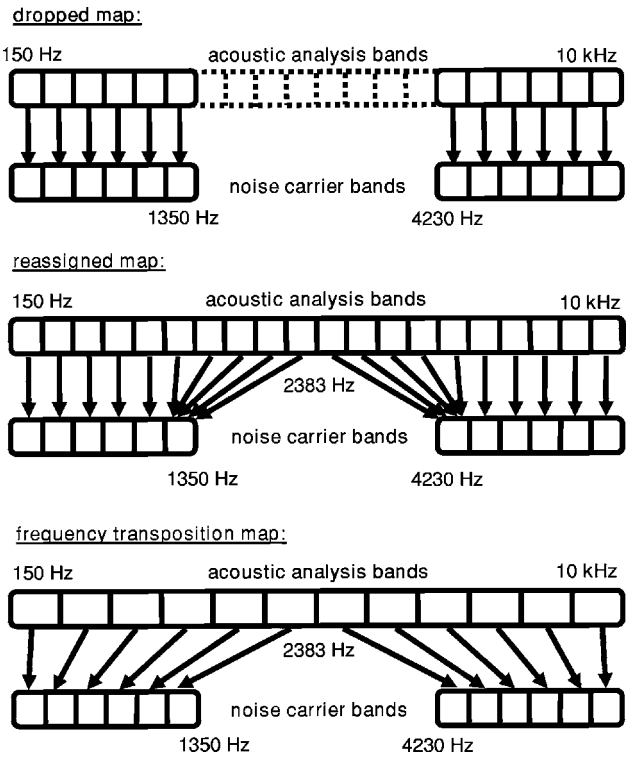


FIG. 1. Three frequency-place mapping conditions with a simulated dead region in the middle cochlear location. In the dropped condition, shown at the top, the acoustic information was removed from the analysis band range, and the remaining frequency range was assigned to matching carrier bands. This map was the first condition in the Shannon *et al.* study and the baseline condition of the present study. In the reassigned condition, shown in the middle, acoustic information was reassigned to the bands neighboring the dead region, simulating the off-place stimulation of auditory nerves. This map was the second condition in the Shannon *et al.* study and was not used in the present study. In the frequency transposition condition, shown at the bottom, the entire input acoustic spectrum was distributed over the remaining carrier bands. This map was used in the present study only.

II. METHOD

A. Subjects

Seven normal-hearing listeners, aged 22 to 36, participated in the experiment. All subjects were native speakers of American English and had hearing thresholds better than 20 dB HL at audiometric frequencies between 125 and 8000 Hz.

The performance of new subjects may increase during the first few testing sessions with noiseband vocoder experiments. The subjects who participated in the present study had

TABLE I. Summary of the simulated dead regions. In columns from left to right, the size of the dead region increases, as shown by the number of the carrier bands dropped. In rows from top to bottom, the simulated cochlear location of the dead region changes from base to apex. The top line in every entry shows which carrier bands were eliminated, numbering starting from the basal end, consistent with the Nucleus electrode array. The corresponding frequency ranges are presented in the bottom lines.

	Size			
Location	2 bands	4 bands	6 bands	8 bands
Basal	5–6	4–7	3–8	2–9
	4230–5635 Hz	3665–6504 Hz	3175–7506 Hz	2571–8664 Hz
Middle	10–11	9–12	8–13	7–14
	2065–2571 Hz	1789–3175 Hz	1550–3665 Hz	1350–4230 Hz
Apical	15–16	14–17	13–18	12–19
	950–1350 Hz	750–1550 Hz	550–1789 Hz	350–2065 Hz

extensive experience with vocoder processing from previous experiments, and they were also familiar with the stimuli. Therefore no training was provided and the data collection started from the first session.

B. Stimuli

The speech recognition tasks consisted of medial vowel and consonant discrimination. Vowel stimuli were taken from materials recorded by Hillenbrand *et al.* (1995). Listeners heard ten presentations, spoken by five male and five female talkers, of 12 medial vowels (ten monophthongs and two diphthongs) presented in a /h/-vowel-/d/ context (heed, hid, head, had, hod, hawed, hood, who'd, hud, heard, hayed, hoed). Chance level on this test was 8.33% correct. Consonant stimuli were taken from materials recorded by Turner *et al.* (1992) and Fu *et al.* (1998). Listeners heard six presentations, spoken by three male and three female talkers, of each of 14 medial consonants /b d f g k m n p s ʃ t θ v z/, presented in an /a/-consonant-/a/ context. Chance level on this test was 7.14% correct.

Subjects were tested once with all 84 consonant presentations and 120 vowel presentations for each condition. All stimuli were presented via loudspeaker in a sound field at 70 dB on an A-weighted scale, without lip-reading. Tokens were presented in random order by custom software (Robert, 1998). Subjects were asked to select the phoneme they heard from a menu displayed on the screen. No feedback was provided. On average, each session lasted from 2 to 4 h, and subjects were able to finish the experiment in three to four sessions.

C. Experimental procedure

A summary of the experimental conditions is given in Table I. From left to right, the size of the simulated dead region increases, and from top to bottom, the location of the simulated dead region moves from base to apex. The top line in every entry shows the number of the bands eliminated from the 20 carrier bands, with the numbering starting from the basal end, consistent with the Nucleus cochlear implant. The corresponding frequency ranges are presented in the bottom lines of each entry.

A noiseband vocoder (Shannon *et al.*, 1995) was implemented in Matlab to simulate the dead regions. To maintain consistency with Shannon *et al.* (2002), the SPEAK Table 9 of Nucleus cochlear implant (Cochlear Corporation, 1995) was used as the basis for the cutoff frequencies of spectral bands. The phonemes were band-pass filtered into 20 bands with sixth-order Butterworth filters, with the overall spectral range of 150 Hz to 10 kHz. The cutoff frequencies of the filters were partitioned linearly for frequencies up to 1550 Hz, and logarithmically for higher frequencies, as shown in the second column of Table II. The speech envelope was extracted from each analysis band by half-wave rectification and low-pass filtering using a third-order Butterworth filter with a cutoff frequency of 160 Hz. The carrier bands were produced by band-pass filtering white noise with sixth-order Butterworth filters. The speech envelopes were used to modulate the noise carrier bands. Eventually, the processed speech stimuli were produced by combining the modulated noise bands.

Dead regions were simulated by eliminating carrier bands in the synthesis stage of the vocoder processing. In the dropped map, analysis bands corresponding to the dead region were discarded, and the remaining analysis bands matched the remaining carrier bands. In the frequency transposition map, the entire analysis frequency range was distributed over the remaining carrier bands. The partitioning of the new analysis bands was similar to the partitioning of the carrier bands. Figure 1 shows an example of a dead region simulated in the middle cochlear location and with different mapping conditions applied. The cutoff frequencies of individual bands are shown in Tables II–IV for all simulated dead region conditions. The second column in each table shows the frequency partitioning of the analysis and carrier bands when there was no dead region. The subsequent columns show the cutoff frequencies of the analysis bands with frequency transposition and dropped maps, as the size of the dead region increases, from left to right. For all dead region conditions, the cutoff frequencies used for the carrier bands with both dropped and frequency transposition maps were the same as the cutoff frequencies used for the analysis bands with the dropped map.

TABLE II. Cutoff frequencies for individual vocoder bands for the dead region simulated in the basal cochlear location. The first column shows the numbering of the bands. The second column shows the partitioning of analysis and carrier bands when there was no dead region. The subsequent columns show cutoff frequencies of analysis bands with the frequency transposition and dropped maps, for the dead regions of 2, 4, 6, and 8 bands. The carrier bands for each dead region condition were the same as the analysis bands with the dropped map.

	None	2 bands, basal		4 bands, basal		6 bands, basal		8 bands, basal	
	analysis and carrier bands (Hz)	frequency transposition (Hz)	dropped map (Hz)	frequency transposition (Hz)	dropped map (Hz)	frequency transposition (Hz)	dropped map (Hz)	frequency transposition (Hz)	dropped map (Hz)
apical end	150	150	150	150	150	150	150	150	150
19-20	350	363	350	379	350	396	350	416	350
18-19	550	581	550	618	550	660	550	710	550
17-18	750	803	750	864	750	937	750	1025	750
16-17	950	1027	950	1117	950	1226	950	1358	950
15-16	1150	1253	1150	1375	1150	1524	1150	1706	1150
14-15	1350	1480	1350	1638	1350	1829	1350	2067	1350
13-14	1550	1710	1550	1904	1550	2142	1550	2440	1550
12-13	1789	1986	1789	2226	1789	2524	1789	2900	1789
11-12	2065	2307	2065	2604	2065	2976	2065	3449	2065
10-11	2383	2679	2383	3046	2383	3509	2383	4102	2383
9-10	2751	3113	2751	3564	2751	4138	2751	4882	2751
8-9	3175	3616	3175	4171	3175	4882	3175	⋯	⋯
7-8	3665	4201	3665	4882	3665	⋯	⋯	⋯	⋯
6-7	4230	4882	4230	⋯	⋯	⋯	⋯	⋯	⋯
5-6	4882	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
4-5	5635	4882	5635	⋯	⋯	⋯	⋯	⋯	⋯
3-4	6504	5842	6504	4882	6504	⋯	⋯	⋯	⋯
2-3	7506	6989	7506	6204	7506	4882	7506	⋯	⋯
1-2	8664	8361	8664	7878	8664	6993	8664	4882	8664
basal end	10 000	10 000	10 000	10 000	10 000	10 000	10 000	10000	10 000

III. RESULTS

The percent correct scores, averaged across subjects, are presented in Figs. 2 and 3 for vowels and consonants, respectively, as a function of the simulated dead region size. The scores were corrected for chance, using the chance levels of 8.33% for vowels and 7.14% for consonants:

$$PC_{\text{corrected for chance}} = (PC_{raw} - \text{Chance level})/(100 - \text{Chance level}).$$

With the adjustment, a raw percent correct score at chance level equals 0% when corrected for chance. In each figure, the panels from left to right present results with the conditions of dead regions simulated in cochlear locations of basal, middle, and apical, respectively. The filled symbols show the scores with the dropped map, and the open symbols show the scores with the frequency transposition map. The error bars show one standard deviation. The dots under the scores indicate the dead region conditions where the performances with the two maps differed significantly, determined by *posthoc* Tukey multiple comparisons. Three dots show a significance level of $p<0.001$, two dots show a significance level of $p<0.01$, and one dot shows a significance level of $p<0.05$.

A two-way repeated measures ANOVA showed that the identification of vowels and consonants decreased significantly with both maps as the size of the dead region increased, however the rate of decrease was different for each map and for each simulated cochlear location (Table V). Similar to the study of Shannon *et al.*, speech recognition was poorer for dead regions simulated in middle and apical locations compared to the basal location. This finding is consistent with the speech intelligibility index (SII) which predicts more contribution from the low- and middle-frequency ranges to overall speech understanding (ANSI S3.5, 1997). The performances observed with the dropped and frequency transposition maps were significantly different, except for the consonant recognition with the conditions of dead regions simulated in middle cochlear locations.

In the dropped map, acoustic information that would fall into the dead region was discarded. The frequency transposition map provided the full input acoustic information, but the mapping of the spectral information to cochlear place was distorted. Frequency transposition was advantageous for consonant recognition with large dead regions (six and eight bands) simulated in middle and apical locations and for vowel recognition with a large dead region simulated in the apical location. However, it produced poorer vowel and consonant recognition with basal dead regions, and poorer vowel

TABLE III. Similar to Table II, except for a dead region simulated in the middle cochlear location.

	None	2 bands, middle		4 bands, middle		6 bands, middle		8 bands, middle	
	analysis and carrier bands (Hz)	frequency transposition (Hz)	dropped map (Hz)	frequency transposition (Hz)	dropped map (Hz)	frequency transposition (Hz)	dropped map (Hz)	frequency transposition (Hz)	dropped map (Hz)
apical end	150	150	150	150	150	150	150	150	150
19-20	350	367	350	388	350	412	350	440	350
18-19	550	591	550	640	550	700	550	772	550
17-18	750	819	750	903	750	1009	750	1137	750
16-17	950	1050	950	1175	950	1333	950	1529	950
15-16	1150	1285	1150	1454	1150	1671	1150	1946	1150
14-15	1350	1521	1350	1739	1350	2022	1350	2383	1350
13-14	1550	1760	1550	2030	1550	2383	1550	⋯	⋯
12-13	1789	2048	1789	2383	1789	⋯	⋯	⋯	⋯
11-12	2065	2383	2065	⋯	⋯	⋯	⋯	⋯	⋯
10-11	2383	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
9-10	2751	2383	2751	⋯	⋯	⋯	⋯	⋯	⋯
8-9	3175	2797	3175	2383	3175	⋯	⋯	⋯	⋯
7-8	3665	3281	3665	2855	3665	2383	3665	⋯	⋯
6-7	4230	3849	4230	3418	4230	2931	4230	2383	4230
5-6	4882	4514	4882	4091	4882	3602	4882	3036	4882
4-5	5635	5293	5635	4894	5635	4422	5635	3860	5635
3-4	6504	6206	6504	5853	6504	5426	6504	4903	6504
2-3	7506	7276	7506	6998	7506	6654	7506	6221	7506
1-2	8664	8530	8664	8366	8664	8158	8664	7889	8664
basal end	10 000	10 000	10 000	10 000	10 000	10 000	10 000	10 000	10 000

recognition with smaller size dead regions in middle and apical locations, compared to the dropped map.

IV. DISCUSSION

The dropped map of the present study was similar to the first condition presented in the Shannon *et al.* (2002) study. The effects of the dropped map on speech recognition observed in the present study were similar to results observed by Shannon *et al.*; there was a decrease in the performance as the size of the simulated dead region increased. The effects were more pronounced for the simulated middle and apical cochlear locations, with the corresponding frequency ranges of 1–2 kHz (Tables III and IV). This finding is consistent with the SII which shows that these frequencies are most important for speech recognition (ANSI S3.5, 1997). Pauler *et al.* (1986) also demonstrated a correlation between poor speech recognition by hearing-impaired subjects and the nerve damage in these cochlear regions. As a result, the decrease in the performance with the dropped map seems to be a direct consequence of missing acoustic information.

In the frequency transposition map, the overall acoustic input range was distributed over the nondead cochlear region. This condition was motivated by the idea of frequency transposition that was previously suggested for hearing aid users with steeply sloping hearing loss (Turner and Hurtig, 1999; McDermott and Dean, 2000) and for cochlear implant users with partial electrode array insertion (Başkent and Shannon, 2005). The frequency transposed condition provides more acoustic information, yet at the expense of introducing distortions in the frequency-place stimulation patterns in the cochlea. Results of the present study showed that the frequency transposition did not necessarily improve speech recognition compared to the dropped map for many simulated dead region conditions; only for six- to eight-band apical dead regions some benefit was observed. This finding cannot be explained by SII, which would predict no drop in performance as the frequency transposition map provides the entire acoustic information for all dead region conditions. The drop in the performance, therefore, is more likely to be caused by the spectral distortions in the mapping of this acoustic information.

Note that the effects observed in the present study are instantaneous effects obtained with normal-hearing subjects. It is possible that listeners might learn over time how to make use of the additional acoustic information, even if presented in a distorted map. Many studies showed a significant improvement in speech recognition by implant users over the course of a few months following the surgery (e.g., Tyler *et al.*, 1997). However, contrary to the high degree of the plasticity shown with pediatric implant users (Sharma *et al.*, 2002), adaptation by adult implant users might be limited. Kral *et al.* (2002), for example, showed a sensitive period of plasticity in the cat auditory system. Başkent and Shannon (2004) measured speech recognition by implant subjects as a

TABLE IV. Similar to Table II, except for a dead region simulated in the apical cochlear location.

	None	2 bands, apical		4 bands, apical		6 bands, apical		8 bands, apical	
	analysis and carrier bands (Hz)	frequency transposition (Hz)	dropped (Hz)	frequency transposition (Hz)	dropped (Hz)	frequency transposition (Hz)	dropped (Hz)	frequency transposition (Hz)	dropped (Hz)
apical end	150	150	150	150	150	150	150	150	150
19-20	350	384	350	443	350	576	350	1150	350
18-19	550	631	550	779	550	1150	550	⋯	⋯
17-18	750	887	750	1150	750	⋯	⋯	⋯	⋯
16-17	950	1150	950	⋯	⋯	⋯	⋯	⋯	⋯
15-16	1150	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
14-15	1350	1150	1350	⋯	⋯	⋯	⋯	⋯	⋯
13-14	1550	1337	1550	1150	1550	⋯	⋯	⋯	⋯
12-13	1789	1563	1789	1362	1789	1150	1789	⋯	⋯
11-12	2065	1826	2065	1612	2065	1383	2065	1150	2065
10-11	2383	2133	2383	1906	2383	1661	2383	1408	2383
9-10	2751	2491	2751	2253	2751	1993	2751	1721	2751
8-9	3175	2908	3175	2662	3175	2389	3175	2099	3175
7-8	3665	3395	3665	3143	3665	2861	3665	2557	3665
6-7	4230	3962	4230	3710	4230	3424	4230	3112	4230
5-6	4882	4624	4882	4378	4882	4096	4882	3785	4882
4-5	5635	5396	5635	5166	5635	4899	5635	4600	5635
3-4	6504	6296	6504	6094	6504	5857	6504	5588	6504
2-3	7506	7356	7506	7189	7506	7002	7506	6786	7506
1-2	8664	8571	8664	8479	8664	8368	8664	8238	8664
basal end	10 000	10 000	10 000	10 000	10 000	10 000	10 000	10 000	10 000

function of the spectral distortions in the frequency-place mapping and analyzed the results for pre- and postlingually deafened subjects. Speech recognition by the prelingually deafened subjects was the best when the frequency-place map of the experimental processor was most similar to the map used in their own implant processor. Performance by the postlingually deafened subjects, however, was best when the experimental map was closest to the normal map of the healthy auditory system. Okazawa *et al.* (1996) similarly observed that postlingually deafened implant users had similar cortical activity to normal-hearing control group, while the PET images of the prelingually deafened subjects were significantly different from both the normal-hearing and the postlingually deafened subjects. These studies suggest that the auditory system of a postlingually deafened subject might be hardwired during the sensitive developmental pe-

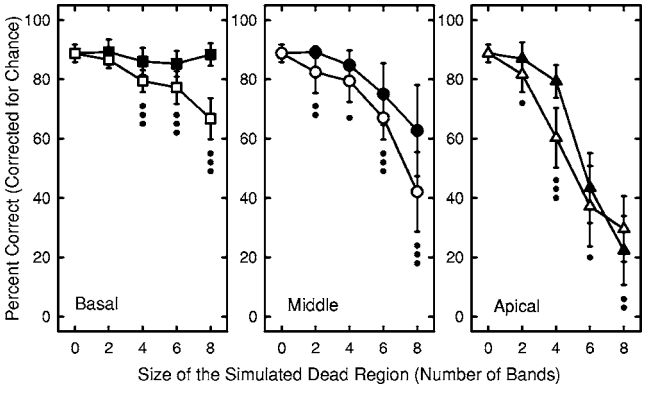


FIG. 2. Percent correct scores, averaged across subjects, for vowel recognition. The scores are presented as a function of the size of the simulated dead region, expressed by the number of carrier bands dropped. The simulated cochlear location of the dead region changes from basal to apical, as shown in the panels from left to right. The filled symbols represent the scores with the dropped map, where the spectrum in the dead region was eliminated from the analysis bands. The open symbols represent the scores with the frequency transposition map, where the entire acoustic input spectrum was distributed over the remaining carrier bands. The error bars show one standard deviation. The dots under the scores show the performances with the two maps that differed significantly, determined by a *posthoc* Tukey test; three dots for $p<0.001$, two dots for $p<0.01$, and one dot for $p<0.05$.

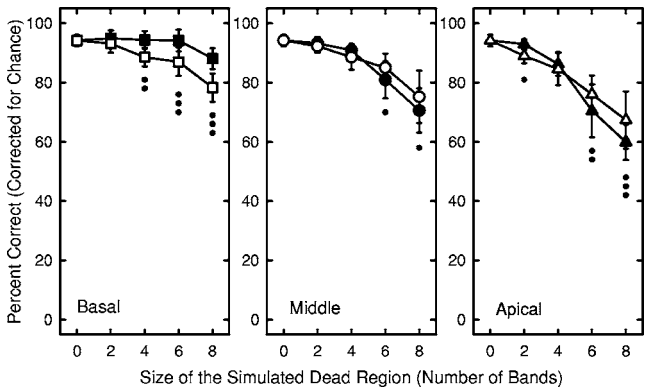


FIG. 3. Similar to Fig. 2, except the speech stimuli were consonants.

TABLE V. F and p values determined by a two-way repeated measures ANOVA. The columns show the values for the dead regions simulated in basal, middle, and apical cochlear locations. The rows show the significance of the effects of the size of the dead region and the mapping condition, and the interaction between the dead region size and the mapping condition, for vowel and consonant recognition.

		Basal	Middle	Apical
Vowel recognition	Size of the dead region	$F(4,24)=35.92$, $p<0.001$	$F(4,24)=47.74$, $p<0.001$	$F(4,24)=143.09$, $p<0.001$
	Mapping condition	$F(1,6)=89.26$, $p<0.001$	$F(1,6)=86.64$, $p<0.001$	$F(1,6)=27.04$, $p<0.01$
	Interaction between the dead region size and the mapping condition	$F(4,24)=23.63$, $p<0.001$	$F(4,24)=12.60$, $p<0.001$	$F(4,24)=13.62$, $p<0.001$
Consonant recognition	Size of the dead region	$F(4,24)=53.15$, $p<0.001$	$F(4,24)=44.51$, $p<0.001$	$F(4,24)=70.16$, $p<0.001$
	Mapping condition	$F(1,6)=34.52$, $p<0.001$	$F(1,6)=1.85$, $p=0.22$	$F(1,6)=8.04$, $p<0.05$
	Interaction between the dead region size and the mapping condition	$F(4,24)=7.22$, $p<0.001$.	$F(4,24)=6.34$, $p<0.01$	$F(4,24)=7.08$, $p<0.001$

riod with normal acoustic patterns, and the adaptation by adult implant users might be limited. Fu *et al.* (2002), for example, introduced a spectral shift in the frequency-place maps of implant users from the normal settings (by 2–4 mm expressed in distance on the cochlea). At the end of the experimental period of 3 months, where the subject had used the shifted map exclusively, speech recognition scores were still lower than the baseline performance obtained with normal settings. Svirsky *et al.* (2004) and Fu *et al.* (2005) showed that adaptation due to simple exposure to a map might be a lengthy process. Fu *et al.* (2005) also showed that a more targeted training might be necessary to facilitate adaptation. Results from the present study can be used as guidelines in fitting implant users. If a matched frequency-place map is more beneficial for the patient in the beginning of the adaptation process, changing the map from the matched map to the frequency transposed map gradually might ease the progress Svirsky *et al.* (2003). In addition, as Başkent and Shannon (2005) showed with simulations of partial electrode array insertions with implant users, there might be an optimal map between the two maps used in the present study, where the frequency transposition would be applied only moderately.

With any simulation study, one should be careful translating the results to real applications with implant users. There are many factors in electrical stimulation that would affect speech perception, such as electrode configuration, stimulus mode, channel interactions, nerve survival pattern, etc. The vocoder processing is only a rough imitation of implant signal processing and does not capture the effects of any of the factors mentioned above. In addition, there are aspects of auditory perception other than speech recognition that might be crucial for implant users, such as the quality of the perceived sounds, which was not covered in the present study.

There is an inherent trade-off in the noiseband vocoder processing; it is usually not possible to preserve the overall acoustic input range, the number of channels, and the bandwidth of individual vocoder bands all at the same time. In the present study, the bandwidth of individual carrier bands was kept the same for all conditions. Therefore, as the size of the simulated dead region increased, the number of channels used in the vocoder decreased. We also had intended to use the entire acoustic input range in the frequency transposition map. When the entire analysis band range was partitioned into a number of channels, the bandwidth of individual analysis bands was larger than the bandwidth of the analysis bands used in the dropped mapping condition. The varying number of channels or bandwidth might have affected the results in addition to the spectral distortions. Follow-up experiments can be designed to complement the present study and to understand the effects of such additional factors. In one experiment the number of channels could be the same for all conditions at the expense of varying the bandwidth of carrier bands and, in another, the bandwidth of all bands could be the same at the expense of losing some acoustic information.

V. CONCLUSION

The results of the present simulation study imply that simply providing more acoustic information to cochlear implant or hearing aid users with dead regions might not be sufficient to improve their speech recognition. If not implemented carefully, spectral distortions introduced by transposing frequencies might cause degradation in speech recognition, especially for postlingually deafened patients. However, there is a possibility that the patients can adapt to the frequency transposition map, especially if training is also provided. Normal-hearing subjects, after all, are used to the redundancy in the speech signal, and the instantaneous effects of spectral distortions might be more detrimental for these listeners compared to real patients, who would have to learn to efficiently use all speech cues available to them. Starting

the patient with a map that is more beneficial for speech recognition and then introducing the frequency transposition in small steps gradually might make such adaptation easier for the patient.

ACKNOWLEDGMENTS

The authors would like to thank Chris Turner, two anonymous reviewers, Emily Olman and Jason Levis, for helpful comments on a previous version of the manuscript, and the subjects for their participation. Funding was provided in part by NIDCD Grant No. R01-DC-01526 and Contract No. N01-DC-92100.

American National Standards Institute (**1997**). "Methods for Calculation of the Speech Intelligibility Index," S35.

Başkent, D., and Shannon, R. V. (**2005**). "Interactions between cochlear implant electrode insertion depth and frequency-place mapping," J. Acoust. Soc. Am. **117**, 1405–1416.

Başkent, D., and Shannon, R. V. (**2004**). "Frequency-place compression and expansion in cochlear implant listeners," J. Acoust. Soc. Am. **116**, 3130–3140.

Bierer, J. A. (2005). "The tripolar electrode configuration and measures of channel interactions in cochlear implant subjects," 28th Midwinter Meeting, Assoc. Res. Otolaryn., New Orleans, LA.

Cochlear Corporation (**1995**). Technical Reference Manual, Englewood, CO.

Florentine, M., and Houtsma, A. J. (**1983**). "Tuning curves and pitch matches in a listener with a unilateral, low-frequency hearing loss," J. Acoust. Soc. Am. **73**, 961–965.

Fu, Q.-J., Nogaki, G., and Galvin, III, J. J. (**2005**). "Auditory training with spectrally shifted speech: Implications for cochlear implant patient auditory rehabilitation," J. Assoc. Res. Otolaryngol. **6**, 180–189.

Fu, Q.-J., Shannon, R. V., and Galvin, III, J. J. (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Fu, Q.-J, Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Gomaa, N. A., Rubinstein, J. T., Lowder, M. W., Tyler, R. S., and Gantz, B. J. (**2003**). "Residual speech perception and cochlear implant performance in postlingually deafened adults," Ear Hear. **24**, 539–544.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Khan, A. M., Whiten, D. M., Nadol, Jr., J. B., and Eddington, D. K. (**2005**). "Histopathology of human cochlear implants: Correlation of psychophysical and anatomical measures," Hear. Res. **205**, 89–93.

Kral, A., Hartmann, R., Tillein, J., Heid, S., and Klinke, R. (**2002**). "Hearing after congenital deafness: Central auditory plasticity and sensory deprivation," Cereb. Cortex **12**, 797–807.

McDermott, H. J., and Dean, M. R. (**2000**). "Speech perception with steeply sloping hearing loss: Effects of frequency transposition," Br. J. Audiol. **34**, 353–361.

Moore, B. C. J., and Glasberg, B. R. (**1997**). "A model of loudness perception applied to cochlear hearing loss," Aud. Neurosci. **3**, 289–311.

Okazawa, H., Naito, Y., Yonekura, Y., Sadato, N., Hirano, S., Nishizawa, S., Magata, Y., Ishizu, K., Tamaki, N., Honjo, I., and Konishi, J. (**1996**). "Cochlear implant efficiency in pre- and postlingually deaf subjects: A study with H2(15)0 and PET," Brain **119**, 1297–1306.

Pauler, M., Schuknecht, H. F., and Thornton, A. R. (**1986**). "Correlative studies of cochlear neuronal loss with speech discrimination and pure-tone thresholds." Arch. Oto-Rhino-Laryngol. **243**, 200–206.

Pfingst, B. E., and Xu, L. (**2004**). "Across-site variation in detection thresholds and maximum comfortable loudness levels for cochlear implants," J. Assoc. Res. Otolaryngol. **5**, 11–24.

Pfingst, B. E., Xu, L., and Thompson, C. S. (**2004**). "Across-site threshold variation in cochlear implants: Relation to speech recognition," Audiol. Neuro-Otol. **9**, 341–352.

Rankovic, C. M. (**1991**). "An application of the articulation index to hearing aid fitting," J. Speech Hear. Res. **34**, 391–402.

Robert, M. E. (**1998**). CONDOR: Documentation for Identification Test Program, House Ear Institute, Los Angeles, CA

Schuknecht, H. F., and Gacek, M. R. (**1993**). "Cochlear pathology in presbycusis," Ann. Otol. Rhinol. Laryngol. **102**, 1–16.

Shannon, R. V., Galvin, III, J. J., and Başkent, D. (**2002**). "Holes in hearing," J. Assoc. Res. Otolaryngol. **3**, 185–199.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Sharma, A., Dorman, M. F., and Spahr, A. J. (**2002**). "Rapid development of cortical auditory evoked potentials after early cochlear implantation," NeuroReport **13**, 1365–1368.

Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., and Mott, J. (**1985**). "Speech perception ability and psychophysical tuning curves in hearing impaired listeners," J. Acoust. Soc. Am. **77**, 620–627.

Svirsky, M. A., Silveira, A., Neuberger, H., Teoh, S. W., and Suarez, H. (**2004**). "Long-term auditory adaptation to a modified peripheral frequency map," Acta Oto-Laryngol. **12**, 381–386.

Svirsky, M. A., Sinha, S., Neuberger, H. S., and Talavage, T. M. (**2003**). "Gradual adaptation to shifts in the peripheral acoustic frequency map," 26th Midwinter Meeting, Asssoc. Res. Otolaryn. Daytona Beach, FL.

Turner, C. W., and Hurtig, R. R. (**1999**). "Proportional frequency compression of speech for listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **106**, 877–886.

Turner, C. W., Fabry, D. A., Barrett, S., and Horwitz, A. R. (**1992**). "Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **35**, 942–949.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., Lowder, M. W., and Gantz, B. J. (**1997**). "Performance over time of adult patients using the Ineraid or Nucleus cochlear implant," J. Acoust. Soc. Am. **102**, 508–522.

Vickers, D. A., Moore, B. C. J., and Baer, T. (**2001**). "Effects of low-pass filtering on the intelligibility of speech in quiet with and without dead regions at high frequencies," J. Acoust. Soc. Am. **110**, 1164–1175.

Autocorrelation in meter induction: The role of accent structure[a)]

Petri Toiviainen and Tuomas Eerola
Department of Music, P.O. Box 35(M), 40014 University of Jyväskylä, Jyväskylä, Finland

(Received 16 April 2005; revised 14 October 2005; accepted 7 November 2005)

The performance of autocorrelation-based meter induction was tested with two large collections of folk melodies, consisting of approximately 13 000 melodies for which the correct meters were available. The performance was measured by the proportion of melodies whose meter was correctly classified by a discriminant function. Furthermore, it was examined whether including different melodic accent types would improve the classification performance. By determining the components of the autocorrelation functions that were significant in the classification it was found that periodicity in note onset locations was the most important cue for the determination of meter. Of the melodic accents included, Thomassen's melodic accent was found to provide the most reliable cues for the determination of meter. The discriminant function analyses suggested that periodicities longer than one measure may provide cues for meter determination that are more reliable than shorter periodicities. Overall, the method predicted notated meter with an accuracy reaching 96% for binary classification and 75% for classification into nine categories of meter. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146084]

PACS number(s): 43.75.Cd [DD] Pages: 1164–1170

I. INTRODUCTION

Most music is organized to contain temporal periodicities that evoke a percept of regularly occurring pulses or beats. The period of the most salient pulse is typically within the range of 400 to 900 ms (Fraisse, 1982; Parncutt, 1994; van Noorden and Moelants, 1999). The perceived pulses are often hierarchically organized and consist of at least two simultaneous levels whose periods have an integer ratio. This gives rise to a percept of regularly alternating strong and weak beats, a phenomenon referred to as meter (Cooper and Meyer, 1960; Lerdahl and Jackendoff, 1983). In Western music, the ratio of the pulse lengths is usually limited to 1:2 (duple meter) and 1:3 (triple meter). Meter in which each beat has three subdivisions, such as 6/8 or 9/8, is referred to as compound meter.

A number of computational models have been developed for the extraction of the basic pulse from music. Modeling of meter perception has, however, received less attention. Large and Kolen (1994) presented a model of meter perception based on resonating oscillators. Toiviainen (1997) presented a model of competing subharmonic oscillators for determining the meter (duple versus triple) from an acoustical representation of music. Brown (1993) proposed a method for determining the meter of musical scores by applying autocorrelation to a temporal function consisting of impulses at each tone onset whose heights are weighted by the respective tone durations. A shortcoming of Brown's (1993) study is that she does not provide any explicit criteria for the determination of meter from the autocorrelation functions.

Although there is evidence that the pitch information present in music may affect the perception of pulse and meter (Dawe *et al.*, 1993; Thomassen, 1982; Hannon *et al.*, 2004), most models of pulse finding developed to date rely only on note onset times and durations. Dixon and Cambouropoulos (2000), however, proposed a multi-agent model for beat tracking that makes use of pitch and amplitude information. They found that including this information when determining the salience of notes significantly improved the performance of their model. Vos *et al.*, (1994) applied autocorrelation to the determination of meter in isochronous or almost isochronous music. They utilized a method similar to that proposed by Brown (1993), except for using the melodic intervals between subsequent tones instead of tone durations as the weighting factor in the autocorrelation analysis. Using a corpus of 30 compositions by J. S. Bach, they found that the maxima of the obtained autocorrelation functions matched the respective bar lengths as indicated in the musical score. As the majority of music is nonisochronous, reliance on mere melodic interval structure in meter induction is probably a special case. In a general case, it can be expected that meter determination be based on both temporal and pitch structure. To address the question of the relative importance of temporal and pitch structure in meter determination, it would be necessary to use a rhythmically more variable set of stimuli.

According to a commonly adopted notion, meter can be inferred from phenomenal, structural, and metrical accents (Lerdahl and Jackendoff, 1983). Phenomenal accents are the primary source of meter and are related to the surface structure of music and arise from changes in duration, pitch, timbre, and dynamics. Pitch-related phenomenal accents, more

[a)]Portions of this work were presented in "The role of accent periodicities in meter induction: a classification study," Proceedings of the 8th International Conference on Music Perception and Cognition, Evanston, IL, August 2004; and "Classification of musical metre with autocorrelation and discriminant functions," Proceedings of the 6th International Conference on Music Information Retrieval, London, 2005.

 0001-4966/2006/119(2)/1164/7/$22.50

commonly referred to as melodic accents, arise from "changes in pitch height, pitch interval or pitch contour" (Huron and Royal, 1996). Using a correlational study, Huron and Royal (1996) investigated the extent to which different types of pitch-related accent correspond to respective metrical positions as indicated by musical notation. Of eight different types of melodic accent, they found that the empirically derived accent by Thomassen (1982) was the only one to correlate significantly with metric position. The lack of correlation of the other accent types however, does, not exclude the possibility that various metric accents exhibit periodic structure that may serve as an additional cue for meter induction. For instance, different types of accents may contain periodicities of equal length, while being phase shifted with respect to each other. This is the case, for instance, with the accents 1, 4, and 5 in Fig. 3. It must be noted that temporal structure and pitch information are not the sole determinants of meter, as it is also influenced by other features such as phrasing and lyrics as well as accents introduced by the performer. These aspects are, nonetheless, beyond the scope of the present study.

A shortcoming of meter induction models presented to date is that they have not been evaluated with large sets of musical material. While Vos *et al.* (1994) utilized a corpus of 30 compositions, Brown (1993) presented only a handful of short musical excerpts to visualize the performance of her model. The present study investigated the performance of autocorrelation-based meter induction with large collections of folk melodies (consisting of thousands of items), for which the notated meters are available. The performance was assessed by the proportion of melodies that were correctly classified in terms of their meter. The components of the autocorrelation function that are significant in the classification were determined. Moreover, a number of different types of melodic accent and combinations thereof were applied to the classification to assess the significance of each of them in the induction of meter. Finally, confusions made by the algorithm between different types of meter were investigated in detail.

II. AUTOCORRELATION AND METER

Below, the method for constructing the autocorrelation function for meter induction is described. For the original description, see Brown (1993). Let the melody consist of N notes with onset times $t_i, i=1,2,\ldots,N$. Each note is associated with an accent value $a_i, i=1,2,\ldots,N$; in Brown (1993), a_i equals the duration of the respective note. The *onset impulse function* f is a time series consisting of impulses of height a_i located at each note onset position:

$$f(n) = \sum_{i=1}^{N} a_i \delta_i(n), \quad n = 0,1,2,\ldots, \tag{1}$$

where

$$\delta_i(n) = \begin{cases} 1, & n = [t_i/dt], \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where dt denotes the sampling interval and [ ] denotes rounding to the nearest integer.

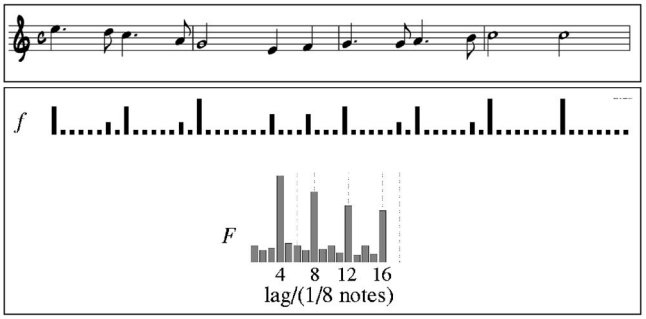


FIG. 1. Excerpt from a melody, its onset impulse function weighted by durational accents, f, and the corresponding autocorrelation function, F. The maximum of the autocorrelation function at the lag of 4/8 indicates duple meter.

Autocorrelation refers to the correlation of two copies of a time series that are temporally shifted with respect to each other. For a given amount of shift (or lag), a high value of autocorrelation suggests that the series contains a periodicity with length equaling the lag. In the present study, the autocorrelation function F was defined as

$$F(m) = \sum_n f(n) f(n-m) \Big/ \sum_n f(n)^2, \tag{3}$$

where m denotes the lag in units of sampling interval; the denominator normalizes the function to $F(0)=1$ irrespective of the length of the sequence. Often, the lag corresponding to the maximum of the autocorrelation function provides an estimate of the meter. This is the case for the melody depicted in Fig. 1.

Sometimes the temporal structure alone is not sufficient for deducing the meter. This holds, for example, for isochronous and temporally highly aperiodic melodies. In such cases, melodic structure may provide cues for the determination of meter. This is the case, for instance, with the melody depicted in Fig. 2. With this isochronous melody, the autocorrelation function obtained from the duration-weighted onset impulse function fails to exhibit any peaks, thus making it impossible to determine the meter. Including information about pitch content in the onset impulse function leads, however, to an autocorrelation function with clearly discernible peaks.

III. STUDY 1

A. Material

The material used in the first study consisted of folk melodies in MIDI file format taken from two collections: the Essen collection (Schaffrath, 1995), consisting of mainly European folk melodies, and the Digital Archive of Finnish Folk Tunes (Eerola and Toiviainen, 2004a), subsequently referred to as the Finnish collection. For the present study, all melodies in either duple (2/4, 4/4, 4/8, etc.; ^2n eighth notes per measure) or triple/compound (3/8, 3/4, 6/8, 9/8, 12/8, etc.; $3n$ eighth notes per measure) meter were chosen. Consequently, a total of 5507 melodies in the Essen collection were used in the study, of which 3121 (56.7%) were in duple and 2386 (43.3%) were in triple/compound meter. From the

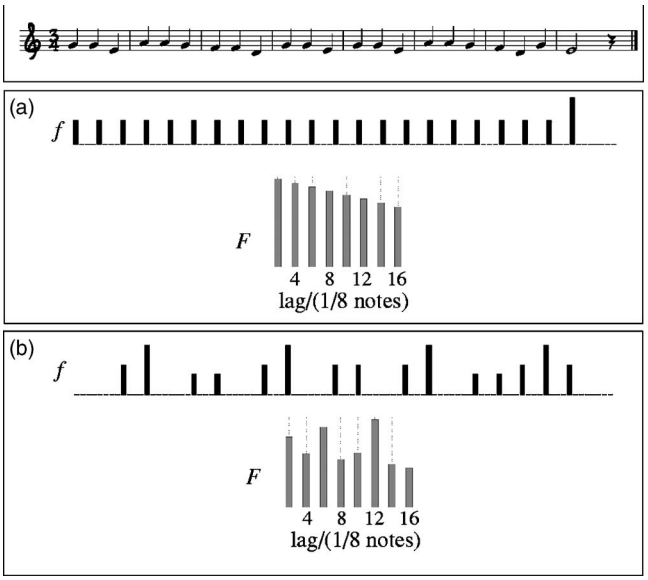


FIG. 2. Excerpt from an isochronous melody. (a) Onset impulse function weighted by durational accents, f, and the corresponding autocorrelation function, F, showing no discernible peaks. (b) Onset impulse function weighted by interval size, f, and the corresponding autocorrelation function, F. The maximum of the autocorrelation function at the lag of 12/8 indicates triple or compound meter.

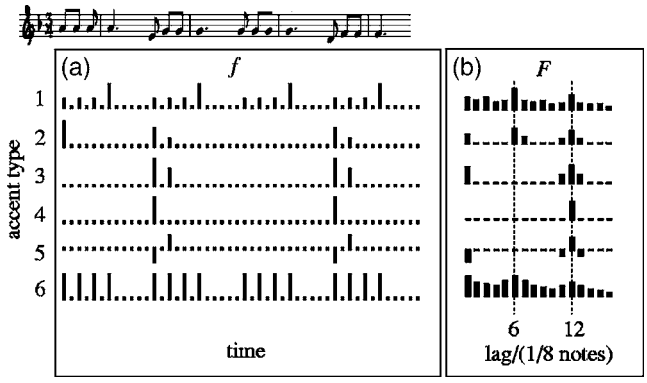


FIG. 3. (a) Onset impulse functions constructed from a melodic excerpt using the six accent types described in the text; (b) the respective autocorrelation functions. As can be seen, the melodic accents frequently fail to co-occur either with each other or with the durational accents. All the autocorrelation functions, however, have maxima at lags of either 6/8 or 12/8, indicating triple or compound meter.

Finnish collection, 6861 melodies were used, of which 5518 (80.4%) were in duple and 1343 (19.6%) were in triple/compound meter.

B. Method

For each of the melodies in the two collections, we constructed a set of onset impulse functions weighted by various accent types [Eqs. (1) and (2)]. In each case the sampling interval was set to 1/16 note. The accents consisted of (1) durational accent (a_i equals tone duration), (2) Thomassen's (1982) melodic accent, (3) interval size in semitones between previous and current tone (see, e.g., Vos *et al.*, 1994), (4) pivotal accent (a_i=1 if melody changes direction, a_i=0 otherwise), and (5) gross contour accent (a_i=1 for ascending interval, a_i=−1 for descending interval, a_i=0 otherwise). Since the note onset times alone, without regard to any accent structure, provide information about metrical structure, we further included (6) constant accent (a_i=1). The analysis was carried out using the MIDI Toolbox for Matlab (Eerola and Toiviainen, 2004b).

For each melody, each of the onset impulse functions was subjected to autocorrelation. The components of the obtained autocorrelation functions corresponding to lags of 1, 2,…, 16 eighth notes were included in the subsequent analyses. Figure 3 depicts the onset impulse functions and the respective autocorrelation functions constructed from a melodic excerpt using each of the accent types described above.

The extent to which these autocorrelation functions could predict the meter of each melody was assessed by means of stepwise discriminant function analyses, in which various subsets of autocorrelation functions were used as independent variables and the meter (duple versus triple/compound) as the dependent variable. The leave-one-out cross-validation scheme (Lachenbruch and Mickey, 1968) was utilized. The measures observed in the analyses were the percentage of correctly classified cases, the order in which variables entered into the discriminant function, and the discriminant function coefficients.

C. Results

The analysis was started by considering the autocorrelation function based on durational accents. First, only the components corresponding to lags of 4 and 6 eighth notes were used as independent variables. This yielded a correct classification rate of 80.9% for the Essen collection and 84.6% for the Finnish collection. Thus, a significant proportion of the melodies was misclassified, suggesting that periodicities of 4/8 and 6/8 in durational accent are not sufficient for reliable classification. This can be clearly seen in Fig. 4, where the values of these components are displayed as scatter plots, showing a significant overlap between melodies representing the two types of meter.

Next, all the components of the durational-accent-based autocorrelation function were entered into the analysis. This yielded a correct classification rate of 90.5% for the Essen collection and 93.1% for the Finnish collection. Inclusion of all 16 components as independent variables thus considerably improved classification performance. The first components to enter into the discriminant function were lags of 8/8, 12/8, and 16/8 (in this order) for the Essen collection, and 4/8, 12/8, and 16/8 for the Finnish collection. This suggests that periodicities longer than one bar may provide cues for meter determination that are more reliable than shorter periodicities.

Subsequently, discriminant function analyses were carried out with the autocorrelation functions obtained using each of the remaining accents, one at a time. The results are summarized in Table I. As can be seen, a significant proportion of the components of the autocorrelation functions that entered first in the stepwise analysis correspond to relatively long time lags. In particular, for all accent types lag 12/8 is among the three most important components. For lag 16/8 the same holds true for all but one accent type. This again suggests that periodicities exceeding the span of one bar

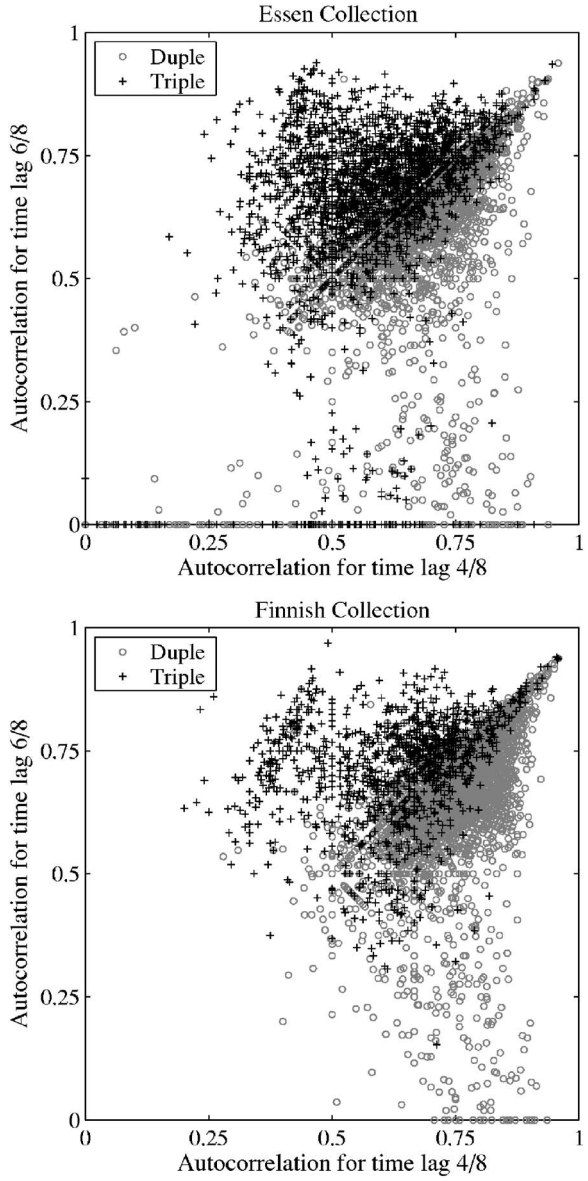


FIG. 4. Scatter plot of the values for lags 4/8 and 6/8 of the durational-accent-based autocorrelation function for the melodies in the Essen collection and the Finnish collection that were used in the present study. Circles and crosses represent melodies in duple and triple meter, respectively.

seem to offer highly important cues for meter induction. Somewhat unexpectedly, the autocorrelation function obtained by ignoring any accent structure (constant accent) yielded for both collections the highest classification rate. In light of this result, the onset function constructed using a constant accent may be slightly more efficient in meter induction than the function originally introduced by Brown (1993). In terms of correct classification rate, the constant accent was followed by durational accent and Thomassen's melodic accent, in this order.

Finally, all autocorrelation functions were used together as independent variables. This yielded a correct classification rate of 95.3% and 96.4% for the Essen collection and the Finnish collection, respectively. The first variable to enter into the discriminant function was lag 8/8 with durational accent for the Essen collection and lag 4/8 with constant accent for the Finnish collection. The next three variables to enter were the same for both collections: lag 12/8 with constant accent, lag 16/8 with constant accent, and lag 6 with melodic accent, in this order. This analysis contained a total of 96 independent variables (16 for each of the 6 accent types); due to the large number of cases, a large number of variables were entered into the stepwise analyses. To obtain a simpler model for meter classification, we performed a further discriminant function analysis in which the six most prominent components for two most prominent accent types from the previous analysis were used. These corresponded to the values of the autocorrelation functions at lags of 3/8, 4/8, 6/8, 8/8, 12/8, and 16/8 using constant accent and Thomassen's melodic accent. Using both collections together (N=12 368), we obtained the following discriminant function,

$$\begin{aligned}\Delta = &-1.042 + 0.318F_{\text{const}}(3/8) + 5.240F_{\text{const}}(4/8)\\ &- 0.630F_{\text{const}}(6/8) + 0.745F_{\text{const}}(8/8)\\ &- 8.122F_{\text{const}}(12/8) + 4.160F_{\text{const}}(16/8)\\ &- 0.978F_{\text{mel}}(3/8) + 1.018F_{\text{mel}}(4/8) - 1.657F_{\text{mel}}(6/8)\\ &+ 1.419F_{\text{mel}}(8/8) - 2.205F_{\text{mel}}(12/8)\\ &+ 1.568F_{\text{mel}}(16/8), \end{aligned} \quad (4)$$

where F_{const} and F_{mel} denote autocorrelation functions ob-

TABLE I. Proportions of correctly classified melodies (classification rate) and components of the autocorrelation function that entered first in each analysis (primary components; numbers refer to lags in units of one eighth note) for both collections and each accent type.

Accent type	(1) Duration	(2) Thomassen	(3) Interval	(4) Pivot	(5) Contour	(6) Constant
			Essen Collection (N=5507)			
Classification rate (%)	90.5	86.8	85.9	78.5	73.1	91.5
Primary Components	8, 12, 16	16, 12, 8	16, 12, 4	16, 12, 6	16, 12, 8	8, 12, 16
			Finnish collection (N=6861)			
Classification rate (%)	93.1	90.1	87.6	79.4	77.9	94.7
Primary Components	4, 12, 16	8, 12, 4	8, 12, 16	6, 16, 12	12, 6, 16	4, 12, 16

TABLE II. Confusion matrices between notated and predicted meters for both collections. Each element shows the number of melodies with the respective notated and predicted meter. For each meter, the recall, precision, and F score values are given.

Essen Collection (N=5592)

		Predicted meter												
		2/4	3/2	3/4	3/8	4/1	4/2	4/4	6/4	6/8	Total	Recall	Precision	F
Notated meter	2/4	1130	0	21	9	0	0	124	0	1	1285	0.88	0.86	0.87
	3/2	0	65	11	0	0	8	16	0	0	100	0.65	0.92	0.76
	3/4	52	0	930	30	0	0	91	106	6	1215	0.77	0.90	0.83
	3/8	23	0	9	168	0	0	0	0	91	291	0.58	0.53	0.55
	4/1	0	1	0	0	36	2	0	0	0	39	0.92	0.75	0.83
	4/2	0	4	0	0	11	148	10	0	0	173	0.86	0.87	0.86
	4/4	98	1	32	0	1	12	1452	1	1	1598	0.91	0.85	0.88
	6/4	0	0	21	0	0	0	9	80	0	110	0.73	0.43	0.54
	6/8	15	0	12	109	0	0	1	1	643	781	0.82	0.87	0.84
	Total	1318	71	1036	316	48	170	1703	188	742				

Finnish Collection (N=*7351*)

		Predicted meter												
		2/4	3/2	3/4	3/8	4/4	5/2	5/4	6/4	6/8	Total	Recall	Precision	F
Notated meter	2/4	2439	17	62	0	740	12	14	5	4	3293	0.74	0.69	0.71
	3/2	5	45	4	0	15	3	0	2	0	74	0.61	0.44	0.51
	3/4	98	23	693	2	16	2	2	65	1	902	0.77	0.86	0.81
	3/8	9	0	17	47	0	0	0	1	55	129	0.36	0.48	0.42
	4/4	958	7	5	0	1203	25	0	5	2	2205	0.55	0.60	0.57
	5/2	0	0	0	0	11	26	1	1	0	39	0.67	0.38	0.49
	5/4	32	6	1	0	0	0	374	0	0	413	0.91	0.95	0.93
	6/4	4	1	9	0	26	0	0	38	0	78	0.49	0.32	0.39
	6/8	14	3	16	48	0	0	1	2	134	218	0.61	0.68	0.65
	Total	3559	102	807	97	2011	68	392	119	196				

tained with constant and Thomassen's melodic accent, respectively. With this discriminant function, the correct classification rates were 92.9% and 94.8% for the Essen and the Finnish collections, respectively. This simpler discriminant function thus yielded correct classification rates that were merely 2.4% and 1.6% lower than those obtained with 96 predictive variables. Variables that received the largest coefficients in the discriminant function were $F_{\text{const}}(12/8)$, $F_{\text{const}}(4/8)$, $F_{\text{const}}(16/8)$, $F_{\text{mel}}(12/8)$, and $F_{\text{mel}}(6/8)$, in this order. In concordance with the aforementioned results, this suggests that temporal structure above the bar level produces important cues for meter determination. Further, the fact that the most significant components of the melodic accent autocorrelation function correspond to multiples of 3/8 lags suggests that, especially for triple and compound meters, melodic accent structure provides additional cues.

IV. STUDY 2

The aim of study 2 was to assess the capability of the autocorrelation-based meter induction method to carry out a more detailed classification. More specifically, instead of mere classification as duple versus triple, the dependent variable used in this study was the actual notated meter. In the analysis, special attention was paid to the pattern of confusion between meters.

A. Material

As in study 1, the material was taken from the Essen collection and the Digital Archive of Finnish Folk Tunes. From each collection, melodies that consisted of a single notated meter were included. Moreover, for each collection only meters that contained more than 30 exemplars were included. Consequently, a total of 5592 melodies in the Essen collection were used, representing nine different notated meters (2/4, 3/2, 3/4, 3/8, 4/1, 4/2, 4/4, 6/4, 6/8). From the Finnish collection, 7351 melodies were used, representing nine different notated meters (2/4, 3/2, 3/4, 3/8, 4/4, 5/2, 5/4, 6/4, 6/8). For each collection, the number of melodies representing each notated meter is shown in Table II.

B. Methods

The classification of meters was based on the discriminant function obtained using the autocorrelation functions obtained using all the accent types. The performance was assessed by means of a confusion matrix. Furthermore, for both collections the precision and recall values as well as the F score were calculated for each meter (Salton and McGill, 1983). For a given meter, precision is defined as the number of melodies having the meter and being correctly classified divided by the total number of melodies being classified as representing the meter. A high value of precision thus indi-

cates that, of the melodies classified as being notated in a given meter, a large proportion is correctly classified. Similarly, for each meter, recall is defined as the number of melodies being notated in the meter and being correctly classified divided by the total number of melodies being notated in the meter. A high value of recall thus indicates that of the melodies being notated in a given meter, a large proportion is correctly classified. The F score is defined as the harmonic mean of precision and recall and is regarded as an overall measure of classification performance (see, e.g., Salton and McGill, 1983).

C. Results

Overall, 83.2% of the melodies from the Essen collection and 68.0% of those from the Finnish collection were correctly classified. These proportions being lower than the corresponding figures in the first study is due to the larger number of classes used in this study (nine) than in the first (two). The notably low correct classification rate for the Finnish collection can be mainly attributed to the fact that a large proportion (43.4%) of the melodies representing 4/4 meter were classified as being 2/4 (see below).

To obtain a more detailed view of the classification performance, we calculated the confusion matrices for both collections. They are displayed in Table II. The table also shows the precision and recall values as well as the F scores for each meter. In terms of the F score, the most accurately classified meters were 4/4 and 2/4 for the Essen collection and 5/4 and 3/4 for the Finnish collection. Similarly, the least accurately classified meters were 6/4 and 3/8 for both collections.

Table II reveals that the most frequent confusions were made within the groups of duple and triple/compound meters, as defined in study 1, whereas confusions across these groups where significantly less frequent. For both collections, meters 2/4 and 4/4 displayed the highest mutual confusion rate, followed by meters 3/4 and 6/4. A large proportion of these misclassifications can probably be attributed to the effect of tempo on the choice of notated meter (cf. London, 2002).

Certain confusions imply more severe misattributions by the algorithm. For instance, 11.7% of the melodies in the Essen collection notated in 3/4 meter were misclassified as representing binary meter (4/4 or 2/4), the corresponding figure for the Finnish collection being 12.6%. In general, duple meters were less frequently misclassified as representing triple/compound meter as vice versa. This asymmetry is in line with the results obtained in study 1. Further research would be needed to account for this phenomenon.

As the confusion matrices contain an abundance of numbers, the relationship between meters may be difficult to see. Therefore we visualized the relations between meters by performing separate hierarchical cluster analyses for both collections. To this end, we calculated the distance between each meter from the confusion matrix according to the formula

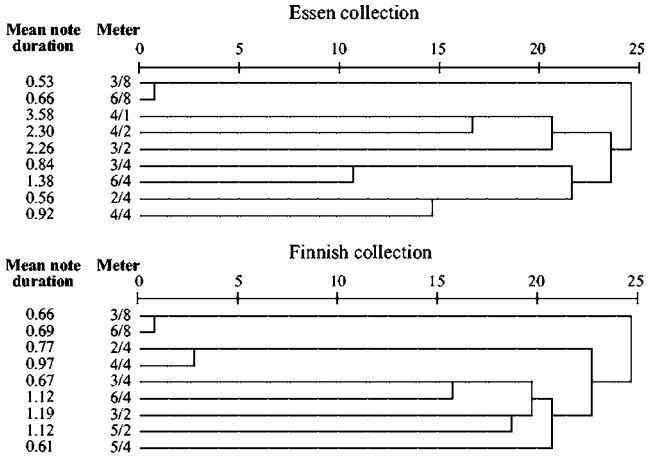


FIG. 5. Dendrograms obtained from the confusion matrix using the similarity measure of Eq. (5). The leftmost column displays the average note durations in quarter notes for the melodies representing each meter.

$$d_{ij} = 1 - \left(\frac{c_{ij} + c_{ji}}{c_{ii} + c_{jj}} \right), \tag{5}$$

where d_{ij} denotes the distance between meters i and j, and c_{ij} is the number of cases where a melody in meter i has been classified as being in meter j. By definition, the larger the proportion of melodies confused between meters, $c_{ij}+c_{ji}$, to the number of melodies correctly classified for both meters, $c_{ii}+c_{jj}$, the smaller the distance d_{ij} between the meters. Figure 5 displays the dendrograms obtained from the clustering algorithms.

In the dendrograms of Fig. 5, the stage at which given meters cluster together reflects the algorithm's rate of confusion between the meters. For both collections, the meters to first cluster together are 3/8 and 6/8. For the Essen collection, this is followed by the clustering of the meters 3/4 and 6/4 as well as 2/4 and 4/4, in this order. Also for the Finnish collection these pairs of meters cluster next, albeit in reverse order, that is, the clustering of 2/4 and 4/4 precedes that of 3/4 and 6/4. A further similar feature between the two dendrograms is that the last clustering occurs between the cluster formed by the meters 3/8 and 6/8 and the cluster formed by all the other meters. This suggests that, in terms of the autocorrelation functions, meters 3/8 and 6/8 are most distinct from the other meters.

One peculiar feature of the dendrogram for the Essen collection is the relatively late clustering of meters 4/1 and 4/2 with meters 2/4 and 4/4. In particular, the former two meters cluster with meter 3/2 before clustering with the latter two meters. A potential explanation for this is the difference in the average note durations between the meters, shown in the leftmost column of Fig. 5. More specifically, the average note durations for meters 4/1, 4/2, and 3/2 exceed those of meters 2/4 and 4/4 by a factor of 2. This anomaly, however, is not significant, as meters 4/1, 4/2, and 3/2 constitute merely a minor proportion of the whole collection.

V. CONCLUSIONS

We studied the classification performance of the autocorrelation-based meter induction model, originally introduced by Brown (1993). Using discriminant function analysis, we provided an explicit method for the classification. Furthermore, we applied the algorithm to investigate the role of melodic accent structure in meter induction.

In conformance with the general view, we found that periodicity in note onset locations was the most important cue for the determination of meter. The results also imply that periodicities longer than one measure provide additional information for meter induction. A somewhat unexpected finding was that ignoring the accent structure in the onset impulse function yielded the best classification rate. This result is difficult to explain and calls for further investigation.

Furthermore, the results suggest that periodicity in melodic accent structure may serve as an additional cue in the determination of meter. In particular, including Thomassen's (1982) melodic accent was found to improve classification performance to some extent. This, however, does not necessarily imply that melodic and durational accents need cooccur but rather that they exhibit similar periodic structure with eventual mutual phase shift.

An additional finding that calls for further study was the significant difference between the correct classification rates for melodies in duple and triple/compound meter. More specifically, melodies in duple meter were more often correctly classified than melodies in triple/compound meter. A detailed investigation of misclassified melodies could provide insight into this question.

The material used in the present study consisted of melodies predominantly in duple, triple, and compound meters, although study 2 utilized a few hundred Finnish melodies in either 5/4 or 5/2 meter. To obtain a deeper insight into the role of accent structure in meter induction, a study with nonregular meters such as those present in the folk music of the Balkan region could be carried out.

An apparent limitation of the method presented in this article is its inability to deal with melodies that contain changes of meter. For a melody that, say, starts in 2/4 meter and changes to 3/4 meter, the algorithm gives unpredictable results. This is due to the fact that the algorithm considers the melody as a whole. The limitation may be overcome by applying a windowed analysis analogical to algorithms used in pitch estimation from acoustical signals, in which the autocorrelation is applied to short windowed segments of the melody, with the window moving gradually throughout the melody.

The present study utilized melodies that were represented in symbolic, temporally quantized form. The choice of stimuli was mainly based on the availability of correct (notated) meters for the melodies in the collections. In principle the method could, however, be applied to performed music in acoustical form as well, at least with a monophonic input. This would require algorithms for onset detection (e.g., Klapuri, 1999), pitch estimation (e.g., Brown and Puckette, 1994; Klapuri, 2003), and beat tracking (e.g., Dixon, 2001; Large and Kolen, 1994; Toiviainen, 1998, 2001).

ACKNOWLEDGMENTS

This work was supported by the Academy of Finland (Project No. 102253).

Brown, J. C. (**1993**). "Determination of meter of musical scores by autocorrelation," J. Acoust. Soc. Am. **94**, 1953–1957.

Brown, J. C., and Puckette, M. S. (**1994**). "A high resolution fundamental frequency determination based on phase changes of the Fourier transform," J. Acoust. Soc. Am. **94**, 662–667.

Cooper, G., and Meyer, L. B. (**1960**). *The Rhythmic Structure of Music* (Univ. of Chicago, Chicago).

Dawe, L. A., Platt, J. R., and Racine, R. J. (**1993**). "Harmonic accents in inference of metrical structure and perception of rhythm patterns," Percept. Psychophys. **54**, 794–807.

Dixon, S. (**2001**). "Automatic extraction of tempo and beat from expressive performances," J. New Music Res. **30**, 39–58.

Dixon, S., and Cambouropoulos, E. (**2000**). "Beat tracking with musical knowledge," in ECAI 2000: Proc. 14th European Conference on Artificial Intelligence (IOS, Amsterdam, Netherlands), pp. 626–630.

Eerola, T., and Toiviainen, P. (**2004a**). *Digital Archive of Finnish Folk Tunes*, University of Jyväskylä, Jyväskylä, Finland. Available at: http://www.jyu.fi/musica/sks/

Eerola, T., and Toiviainen, P. (**2004b**). *MIDI Toolbox: MATLAB Tools for Music Research*, University of Jyväskylä, Jyväskylä, Finland. Available at http://www.jyu.fi/musica/miditoolbox/

Fraisse, P. (**1982**). "Rhythm and tempo," in *Psychology of Music*, edited by D. Deutsch (Academic, New York), pp. 149–180.

Hannon, E., Snyder, J., Eerola, T., and Krumhansl, C. L. (**2004**). "The Role of melodic and temporal cues in perceiving musical meter," J. Exp. Psychol. Hum. Percept. Perform. **30**, 956–974.

Huron, D., and Royal, M. (**1996**). "What is melodic accent? Converging evidence from musical practice," Music Percept. **13**, 489–516.

Klapuri, A. (**1999**). "Sound Onset Detection by Applying Psychoacoustic Knowledge," Proc. IEEE Int. Conf. Acoustics Speech and Sig. Proc. (ICASSP), Phoenix, AZ, pp. 3089–3092.

Klapuri, A. (**2003**). "Multiple fundamental frequency estimation by harmonicity and spectral smoothness," IEEE Trans. Speech Audio Process. **11**, 804–816.

Lachenbruch, P. A., and Mickey, M. R. (**1968**). "Estimation of error rates in discriminant analysis," Technometrics **10**, 1–11.

Large, E. W., and Kolen, J. F. (**1994**). "Resonance and the perception of musical meter," Connection Sci. **6**, 177–208.

Lerdahl, F., and Jackendoff, R. (**1983**). *A Generative Theory of Tonal Music* (MIT, Cambridge, MA).

London, J. (**2002**). "Cognitive constraints on metric systems: some observations and hypotheses," Music Percept. **19**, 529–550.

Parncutt, R. (**1994**). "A perceptual model of pulse salience and metrical accent in musical rhythms," Music Percept. **11**, 409–464.

Salton, G., and McGill, M. (**1983**). *Introduction to Modern Information Retrieval* (McGraw Hill, New York).

Schaffrath, H. (**1995**). *The Essen Folksong Collection in Kern Format* (computer database), edited by D. Huron (Center for Computer Assisted Research in the Humanities, Menlo Park, CA).

Thomassen, J. M. (**1982**). "Melodic accent: Experiments and a tentative model," J. Acoust. Soc. Am. **71**, 1596–1605.

Toiviainen, P. (**1997**). "Modelling the perception of metre with competing subharmonic oscillators," Proc. Third Triennial ESCOM Conference, Uppsala, Uppsala University, pp. 511–516.

Toiviainen, P. (**1998**). "An interactive MIDI accompanist," Comput. Music J. **22**, 63–75.

Toiviainen, P. (**2001**). "Real-time recognition of improvisations with adaptive oscillators and a recursive Bayesian classifier," J. New Music Res. **30**, 137–148.

van Noorden, L., and Moelants, D. (**1999**). "Resonance in the perception of musical pulse," J. New Music Res. **28**, 43–66.

Vos, P. G., van Dijk, A., and Schomaker, L. (**1994**). "Melodic cues for metre," Perception **23**, 965–976.

Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates

Bruno L. Giordano[a),b)]
Dipartimento di Psicologia Generale, Università degli Studi di Padova, Via Venezia 8, I-35100 Padova, Italy and Institut de Recherche et Coordination Acoustique/Musique (STMS-IRCAM-CNRS), 1 Place Igor Stravinsky, F-75004 Paris, France

Stephen McAdams[a),c)]
Département d'Etudes Cognitives, Ecole Normale Supérieure, 45, Rue d'Ulm, F-75230 Paris Cedex 05, France and STMS-IRCAM-CNRS, 1 Place Igor Stravinsky, F-75004 Paris, France

(Received 22 March 2005; accepted 14 November 2005)

Identification of the material of struck objects of variable size was investigated. Previous studies on this issue assumed recognition to be based on acoustical measures of damping. This assumption was tested, comparing the power of a damping measure in explaining identification data with that of several other acoustical descriptors. Listeners' performance was perfect with respect to gross material categories (steel-glass and wood-plexiglass) comprising materials of vastly different mechanical properties. Impaired performance was observed for materials within the same gross category, identification being based on the size of the objects alone. The damping descriptor accounted for the identification of the gross categories. However other descriptors such as signal duration explained the results equally well. Materials within the same gross category were identified mainly on the basis of signal frequency. Overall poor support for the relevance of damping to material perception was found. An analysis of the acoustical support for perfect material identification was carried out. Sufficient acoustical information for perfect performance was found. Thus, procedural biases for the origin of the effects of size could be discarded, pointing toward their cognitive, rather than methodological nature. Identification performance was explained in terms of the regularities of the everyday acoustical environment. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2149839]

PACS number(s): 43.75.Cd, 43.66.Jh, 43.75.Kk [NHF] Pages: 1171–1181

I. INTRODUCTION

A growing branch of research, variously labelled *ecological acoustics* (Vanderveer, 1979), *auditive kinetics* (Guski, 2000), *psychomechanics* (McAdams, 2000), or, generally, sound source perception, investigates the perceptual correlates of the properties of sound sources. The object of study in this field can be described at three different levels: physical or mechanical (the properties of the sound source), acoustical (the properties of the sound wave emitted by the source), and perceptual (the perceived properties of the sound event). The research design in sound source recognition analyzes all the pairwise relationships among these levels (Li, Logan, and Pastore, 1991). In the present study this pairwise design was applied to one of the most investigated issues, identification of material type in impact sounds, making it possible to provide a structured framework for the understanding of everyday perception of this source property.

The vast majority of previous studies on material identification focused on the effects of acoustical measures of damping. Wildes and Richards (1988) defined a shape invariant acoustical parameter for material type, the coefficient of internal friction $\tan\phi$, which models material damping,

$$\tan\phi = \frac{\alpha}{\pi f}, \tag{1}$$

where α is the damping coefficient of the vibrational component, i.e., the inverse of the time required for vibration amplitude to decay to $1/e$ of its original amplitude, and f is its frequency. The higher $\tan\phi$ the greater the damping of the material and the faster the decay time decreases with increasing frequency. Wildes and Richards (1988) proposed material type recognition to be based on the $\tan\phi$ coefficient.

The effects of damping measures on material identification were tested in several studies using both synthetic and real sounds. Klatzky, Pai, and Krotkov (2000) investigated stimuli synthesized according to a physical model of a struck bar (van den Doel and Pai, 1998), varying a parameter related to $\tan\phi$ and the frequency of the lowest vibrational mode, later referred to as frequency, which spanned over 3.3 octaves. Four response categories were used: rubber, wood, steel, and glass. Both experimental variables affected identification: rubber and wood were chosen for higher $\tan\phi$ values than glass and steel; glass and wood were chosen for higher frequencies than steel and rubber. The same task was adopted by Avanzini and Rocchesso (2001). Stimuli were

[a)]Current address: CIRMMT, Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montréal, QC, Canada H3A 1E3.
[b)]Electronic mail: bruno.giordano@music.mcgill.ca
[c)]Electronic mail: smc@music.mcgill.ca

generated according to the physical model of a one-mode resonator, varying the $\tan\phi$ coefficient and frequency (range: 1 octave). Results were analogous to those of Klatzky *et al.* (2000), although frequency effects were less clear. Roussarie (1999) synthesized stimuli according to a physical model of a struck plate (Lambourg, Chaigne and Matignon 2001), varying damping coefficients, elastic properties, and density of the simulated plates around those characterizing glass and aluminum. The properties of the simulated hammer were also manipulated, using parameters typical of either wood or rubber. Two response categories were adopted: glass and aluminum. Identification was influenced only by the damping properties of the plates, strongly correlated with an acoustical parameter analogous to $\tan\phi$ and with the average spectral center of gravity. Variations in density and elasticity, associated with a frequency variation equivalent to a musical interval of a perfect fifth, had no effect. In summary, all these studies demonstrated material identification to be influenced by damping measures, while frequency was relevant only when ranging over at least one octave.

Other studies focused on material identification performance. Gaver (1988) studied variable-length bars made of iron or wood. High recognition performance was observed and bar length had no effect. Kunkler-Peck and Turvey (2000) investigated variously shaped plates made of steel, wood or plexiglass. Performance was almost perfect with only a secondary tendency to associate materials with shapes. Perfect performance was not confirmed, however, in a study conducted on synthetic signals (Lutfi and Oh, 1997). Stimuli were synthesized according to the wave equation of a struck clamped bar, with stimulus variability created by perturbing the density and elasticity terms. Participants were asked which of two stimuli was generated by striking a given target material (iron or glass), the alternatives being different metals, crystal or quartz. Signal frequency was given a disproportionate weight by listeners, resulting in poor performance.

Inconsistencies between results by Kunkler-Peck and Turvey (2000) and by Lutfi and Oh (1997) were explained by Carello, Wagman, and Turvey (2003) in terms of the lack of acoustical richness that might characterize synthetic signals, and thus of the absence of sufficient information for the task. However, additional studies also found impaired performance with real signals. Giordano (2003) studied rectangular steel, glass, wood, and plexiglass plates. Different stimulus sets were generated, varying also the height/width ratio of the plates and their area (both with freely vibrating and externally damped plates), as well as the material of the hammer. With freely vibrating plates identification was perfect only with respect to two gross material categories (wood-plexiglass and glass-steel), strong confusions being found within the categories. Also, externally damped glass plates were identified as made of wood or plexiglass. In any case, consistently with results of Gaver (1988) and Kunkler-Peck and Turvey (2000), identification of gross categories was not influenced by the geometrical properties of the objects. Further, the height/width ratio and hammer material variables had no significant effect. Results by Giordano (2003) were confirmed by Tucker and Brown (2003) with stimuli generated by striking variably shaped wood, plexiglass, and aluminum plates both in open air and underwater. Wood and plexiglass were strongly confused with one another and were almost perfectly discriminated from steel. A parameter related to $\tan\phi$ explained a large portion of the data variance (62–69%). In summary, Lutfi and Oh (1997), Giordano (2003), and Tucker and Brown (2003) found that recognition abilities were limited and were perfect only when involving comparisons among materials of vastly different properties (e.g., woods and metals).

Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003) found the wood/plexiglass identification to be independent of the geometrical properties of the objects. In contrast, Giordano (2003) found identification within both the above-mentioned gross material categories to be based on plate area (larger plates being more frequently identified as made of steel or plexiglass) both with freely vibrating and externally damped plates. These results appeared strongly consistent across listeners, although a small percentage of participants associated wood/plexiglass with large/small plates. Given the influence of plate geometry on signal frequency, the effects of area on identification would seem to confirm results by Klatzky *et al.* (2000). The informal nature of the acoustical analyses presented in Giordano (2003), however, does not allow us to draw conclusions at this point. Indeed participants might have based their judgment on acoustical parameters other than frequency that are also affected by size variations. Furthermore, the effects of size observed by Giordano (2003) might have been caused by the absence of acoustical information which reliably discriminated between materials within the gross categories. Indeed, in the absence of such information, participants might have been forced to focus on source properties irrelevant to the task, namely size, for which a variation in the acoustical features was present. However, no test for the presence of sufficient acoustical information for perfect material identification was carried out.

Despite all the studies focusing on this topic, little is known about the acoustical criteria for material identification, because acoustical modeling of behavioral data was based on limited sets of descriptors including, at best, an acoustical measure of damping, frequency, and the average spectral center of gravity (SCG) (Roussarie, 1999). Furthermore, the ascertained association of judgments with acoustical measures of damping is not sufficient to conclude as to their relevance to material identification, where judgments might instead be based on correlated signal properties like duration, which is expected to increase with decreasing $\tan\phi$. This hypothesis is at least in part supported by the results of a recent study conducted on synthetic struck bar signals (McAdams, Chaigne, and Roussarie, 2004), based on a judgment shown in other studies (see Grey, 1977; and McAdams, 1993) to be strongly related to identification. Consistently, McAdams *et al.* (2004) found dissimilarity of impact sounds to be influenced by level-decay-rate and SCG-related descriptors, both covarying with a measure of the damping in the simulated bars.

A new study on material identification was performed, using a subset of the real signals investigated by Giordano

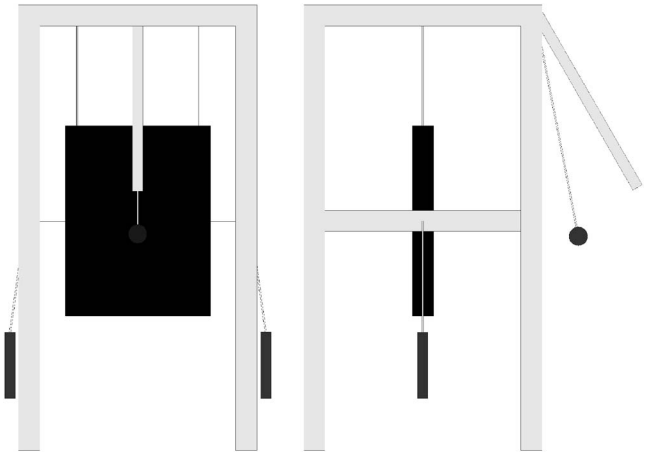


FIG. 1. Sketch of the device used to suspend and strike the plates. The pendulum and the stabilizing weights are shown in dark grey.

(2003). The complete research design in sound source recognition was adopted (Li *et al.*, 1991). A wide set of descriptors was used to characterize the information for identification available to the perceiver, extracting the vast majority of them from a simulation of the basic properties of the peripheral auditory system. For simplicity this level of characterization was termed "acoustical." Both the mechanical and acoustical determinants of experimental judgments were outlined. In particular, the power of several acoustical descriptors in explaining material identification was compared with that of a descriptor closely related to the $\tan\phi$ measure of damping. An analysis of the relationship between the acoustical and physical levels allowed us to test for the presence of sufficient acoustical information for perfect material recognition and, consequently, to test for a procedural origin of the effects of size observed by Giordano (2003). Given the interindividual differences for the wood and plexiglass recognition strategies reported by Giordano (2003), interparticipant agreement was also studied.

II. METHODS

A. Stimuli

Sounds were generated striking 2-mm-thick square plates made of four different materials: plexiglass (polymethyl methacrylate), soda-lime glass, steel, and Tanganyka walnut. Five different values were used for the length of the sides of the plates: 8.66, 12.24, 17.32, 24.49, and 34.64 cm, yielding areas from 75 to 1200 cm^2. Each plate was drilled close to the right and left top corners and close to the left and right borders, at the middle of their height (diameter, 4 mm). The upper holes were used to suspend the plates; the lower ones to stabilize them after being struck, thus avoiding amplitude modulations due to an excessive movement of the plate relative to the microphone. Plates were struck with a steel pendulum (diameter, 2 cm; weight, 35.72 g).

The apparatus used to suspend the plates was similar to that used by Kunkler-Peck and Turvey (2000) (see Fig. 1) and was made of pine wood. Both the plates and the pendulum were hung from the top shelf with nylon lines (diameter, 1 mm). The lateral holes of the plates were attached to two 150-g weights with nylon lines, passing through holes drilled in two horizontal planks attached to both sides of the structure. The pendulum was hung from the top shelf, 15 cm from the plane in which the plates lay, and was released from a fixed guide attached to the front of the top shelf, thus keeping constant the starting angle. Plates were struck in their centers. No audible multiple impacts of the pendulum on the plate were observed.

Sounds were generated in an acoustically isolated room with highly absorbing walls and were recorded using a TASCAM DA-P1 DAT recorder (48000-Hz sampling rate, 16 -bit resolution) and Beyer Dynamic digital microphone (MCD101/MPD200) positioned 45 cm from the center of the plate, opposite the struck surface. Recordings were transferred to a computer hard disk through the digital input of a Sound Blaster Live Platinum sound card. Signals longer than 1 s were reduced to this duration by applying a 5-ms linear decay. Informal listening tests showed that material identification was not influenced by this sound wave editing process. Signals were not equalized in loudness. The presentation level was the maximum level which kept the background noise, constant across the samples, inaudible. The peak levels of the signals ranged from 54 to 72 dB SPL.

B. Procedure

Stimuli where presented through AKG K240 headphones, connected to a Nikko NA-690 amplifier, which received the output of the Sound Blaster Live soundcard of the PC used to program the experiment. Participants sat inside a soundproof booth. They were told that on each trial they would be presented a single sound generated by striking an object made of one of four different materials. In order to make instructions straightforward, it was decided to use generic linguistic labels for all materials: glass for soda-lime glass, metal for steel, plastic for plexiglass, and wood for Tanganyka walnut. As the stimulus set comprised only one material type per generic category, it was assumed that this linguistic choice would not affect participants' responses. No mention was made of the geometrical properties of the objects, minimizing the non-auditory information given to participants. After presentation of the stimulus, participants were asked to identify the material of the struck object. Confidence with the response was preferred to the control of participants exposure to stimuli and of the response time, in order to provide the best possible conditions for the use of the auditory information carried by the stimuli. Thus, before giving the response, participants were allowed to replay the stimulus as many times as needed and were given no constraints on the time required to emit the response. Responses were given by pressing appropriate keyboard keys. The 20 stimuli were presented in block-randomized order for each of seven repetitions, for a total of 140 trials.

C. Participants

Twenty-five listeners took part in the experiment on a voluntary basis (age: 22–49 years; 17 males, 8 females). Given the absence of reported effects on source perception

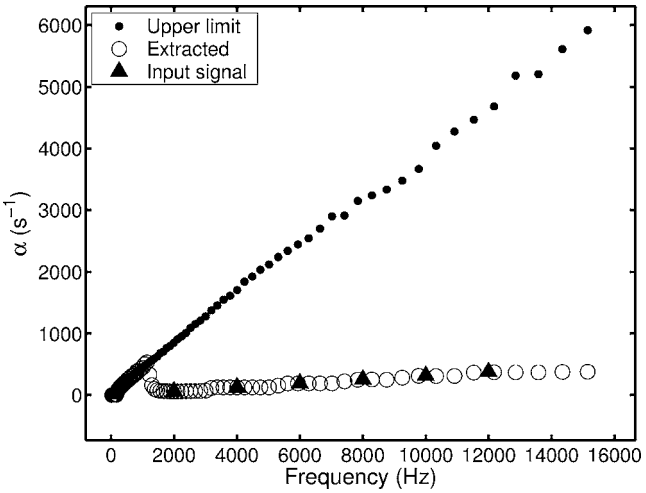


FIG. 2. Damping factors extracted from a six-component harmonic complex with fundamental frequency of 2000 Hz and unweighted tanϕ of 0.01.

performance (see McAdams, 1993, for a review of this literature), gender was not controlled. All of the participants reported having normal hearing.

III. ACOUSTICAL DESCRIPTORS

The analysis model used to extract the vast majority of acoustical descriptors was meant to simulate the output of the cochlea in response to the incoming acoustical signal. Outer and middle ear filtering were simulated with a cascade of two IIR and one FIR filters, in order to account for peak sensitivity at 2 kHz and for loss of sensitivity at lower and higher frequencies. The transfer function was derived from measures of the minimum audible field (Killion, 1978). Processing of the signal inside the cochlea was simulated with a gammatone filter bank (Patterson, Allerhand, and Giguère, 1995), with center frequencies f_c uniformly spaced on an equal-resolution scale (Moore and Glasberg, 1983) between 30 and 16000 Hz. The power in output from the cochlear filters was then added to the power delayed by $1/4f_c$ (Marozeau, de Cheveigné, McAdams, and Winsberg, 2003).

A parameter analogous to tanϕ was extracted from this representation and, given the focus of the analysis model on the properties of the peripheral auditory system, termed tanϕ_{aud}. Damping factors α for the signal output from each channel were computed using the regression model $\log(P) = a + bT$, where P is power, T is time, and $b = -\alpha/2$. The regression model was applied to the signal from peak power to a fixed threshold power. Figure 2 shows the analysis of a harmonic complex given by the sum of six damped sinusoids with a fundamental frequency of 2000 Hz, and with damping factors chosen to yield a tan ϕ (unweighted) of 0.01 [see Eq. (1)]. Also shown is the upper limit for the damping factor of the signal in output from the cochlear channels, calculated analyzing a unitary amplitude impulse.

Tanϕ_{aud} was computed from the damping factors weighting for the total power in output from the cochlear filters, where the higher the output power, the higher their perceptual relevance and thus weight in determining the value of this descriptor. Thus tanϕ_{aud} was defined as

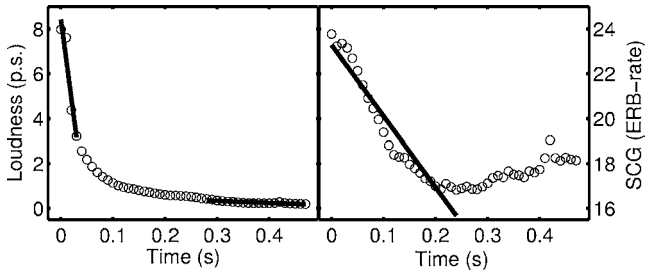


FIG. 3. Temporal functions of loudness and SCG for the signal generated by striking the 150 cm^2 glass plate. Also shown are the linear regression functions used to extract the slope measures $\text{Lou}_{\text{sl1}}, \text{Lou}_{\text{sl2}}, \text{SCG}_{\text{slo}}$.

$$\frac{\sum_{i=1}^{N} \frac{\alpha_i}{\pi f_{ci}} w_i}{\sum_{i=1}^{N} w_i}, \tag{2}$$

where f_{ci} is the center frequency in Hertz, and w_i is the sum of power from peak value to threshold. This procedure yielded, for the signal shown in Fig. 2, a tanϕ_{aud} of 0.0101.

A second representation was used to extract loudness- and brightness-related descriptors. The representation used to compute the tanϕ_{aud} parameter was downsampled, convolving it with a 10-ms square window, yielding a temporal resolution similar to that for loudness integration (Plack and Moore, 1990). The power in each channel was finally raised to the power of 0.25 to approximate partial loudness (Hartmann, 1997). For each temporal frame of this representation, loudness and brightness were defined, and computed, as the sum of the partial loudnesses (Zwicker and Fastl, 1999) and as the spectral center of gravity (SCG) (specific loudness weighted average of frequency), respectively. Finally, a duration (Dur) measure was extracted, offset loudness being that of the background noise (about 0.2 pseudosones).

Attack and average values were extracted from the temporal functions of loudness and SCG ($\text{Lou}_{\text{att}}, \text{SCG}_{\text{att}}$; $\text{Lou}_{\text{mea}}, \text{SCG}_{\text{mea}}$). For 17 of the sounds, the SCG_{att} measure corresponded to the maximum SCG value, while for the remaining three signals, the peak was found in the third analysis frame (20–30 ms from onset). With loudness, the attack corresponded to maximum loudness in nine signals, while for the remaining 11 maximum loudness was found in the second analysis frame (10–20 ms from onset). Further descriptors characterized the temporal evolution of these measures and were extracted with linear regression. For loudness, Lou_{sl1} measured the slope from the attack to the point where loudness reached half of the attack value; Lou_{sl2} measured the slope from the point were loudness was double the final value up to the end. The SCG-over-time function was nonmonotonic for 15 signals, for which an initial decrease was followed by a final increase. Only one slope was extracted (SCG_{slo}), taking into account the portion from attack to the minimum value. Figure 3 shows the loudness and SCG functions over time for the signal generated by striking the 150-cm^2 glass plate. Also shown are the linear regression functions used to extract the slope measures.

Finally, a measure of the frequency of the lowest spectral component F was extracted, on the basis of the fast Fou-

TABLE I. Acoustical descriptors extracted from each signal. Mat.=material; S=steel; G=glass, W=wood; P=plexiglass; ρ=density; p.s.=pseudo-sones. See text for an explanation of the meaning of each acoustical descriptor.

Mat.	Area (cm^2)	ρ (kg/m^3)	$\tan\phi_{aud}\times10^{-3}$	Dur (s)	F (Hz)	Lou_{att} (p.s.)	Lou_{mea} (p.s.)	Lou_{sl1} (p.s./s)	Lou_{sl2} (p.s./s)	SCG_{att} (ERB-rate)	SCG_{mea} (ERB-rate)	SCG_{slo} (ERB-rate/s)
S	75	7708.30	1.05	0.98	1535.15	7.24	0.95	−138.84	−0.62	25.98	21.41	−6.00
S	150	7708.30	0.86	0.98	773.44	6.89	1.45	−45.61	−0.96	24.41	19.90	−5.90
S	300	7708.30	0.90	0.98	386.72	6.17	1.81	−25.04	−1.24	23.27	20.64	−2.85
S	600	7708.30	0.37	0.98	187.50	6.80	3.15	−11.96	−2.50	24.10	21.69	−2.63
S	1200	7708.30	0.27	0.98	93.75	5.84	3.88	−4.78	−2.93	23.83	20.35	−2.97
G	75	2301.70	1.52	0.52	1406.25	8.09	1.25	−153.39	−1.17	25.38	22.48	−7.70
G	150	2301.70	4.46	0.47	750.00	7.97	1.09	−175.29	−0.93	23.76	18.59	−31.70
G	300	2301.70	2.59	0.63	386.72	8.58	1.50	−105.47	−1.12	23.07	19.33	−5.72
G	600	2301.70	1.68	0.98	187.50	7.06	1.47	−42.14	−0.69	22.96	16.86	−7.26
G	1200	2301.70	2.55	0.94	105.47	6.59	1.34	−38.41	−0.54	22.56	17.20	−5.56
W	75	718.33	19.29	0.17	527.34	5.19	0.93	−175.51	−1.40	23.51	18.18	−171.61
W	150	718.33	22.33	0.19	257.81	4.55	0.95	−131.95	−2.03	22.34	16.44	−102.38
W	300	718.33	19.03	0.30	128.91	4.56	0.83	−121.13	−1.12	21.40	15.75	−44.01
W	600	718.33	19.78	0.16	58.60	4.15	1.07	−104.69	−3.41	20.98	16.69	−52.39
W	1200	718.33	17.55	0.23	23.44	4.06	1.05	−64.57	−2.64	21.00	16.10	−36.04
P	75	1413.30	26.09	0.10	527.34	4.78	1.11	−176.74	−3.99	23.26	18.36	−153.04
P	150	1413.30	39.62	0.10	281.25	4.05	1.15	−127.72	−4.52	22.11	17.25	−148.12
P	300	1413.30	41.03	0.13	140.63	3.83	1.10	−110.59	−3.19	21.33	16.84	−114.28
P	600	1413.30	31.03	0.16	70.31	3.71	0.91	−99.08	−2.50	20.98	16.46	−123.87
P	1200	1413.30	24.50	0.17	35.16	3.79	0.91	−84.09	−2.46	20.80	15.38	−85.39

rier transform of the first 4096 samples of the signals (Hanning window). F was defined as the frequency of the first amplitude peak exceeding a fixed threshold. Amplitude threshold was defined as the maximum amplitude of the low-frequency background noise across the recorded samples. Table I shows for each signal the extracted acoustical indices. Approximate density measures for the investigated materials are also given. Notably, $\tan\phi_{aud}$ discriminated perfectly among material types, this measure increasing from steel to glass to wood to plexiglass.

IV. RESULTS

Due to the repetitions, for each sound a distribution of responses across the four categories was possible for each listener. Analyses were conducted on the individual modes of these distributions, hereafter referred to as "modal responses."

A. From physics to perception

Response profiles of small groups of participants presented macroscopic differences with respect to data pooled across all participants. Cluster analysis was used to extract groups of homogeneous response profiles. Distances among individuals were calculated using a general nominal dissimilarity measure, defined as the proportion of consistent categorizations among two participants (Gordon, 1999). An agglomerative hierarchical algorithm (average linkage) was used. The final number of clusters was chosen considering a set of statistical indices that measure the goodness-of-fit between the input data and the resulting clustering partitions (Milligan, 1996). A subset of the available indices was chosen that demonstrates superior performance in recovering the correct number of clusters (Milligan, 1981; Milligan and Cooper, 1985): the c index (Hubert and Levin, 1976), the Goodman-Kruskal γ (Baker and Hubert, 1972), and the point biserial correlation (Milligan, 1980). For the first index, lower scores indicate higher goodness-of-fit, and better partitions; for the latter two higher scores characterize better partitions. Following the suggestions by Gordon (1999), the number of clusters was chosen considering local maxima/minima of these indices across partition levels, the correct number of clusters being characterized by the highest concordance among indices. Figure 4 shows the value of the three indices as a function of the number of clusters, along with the local maxima/minima.

The final number of clusters was taken to be equal to three, this partitioning level being indicated by all three in-

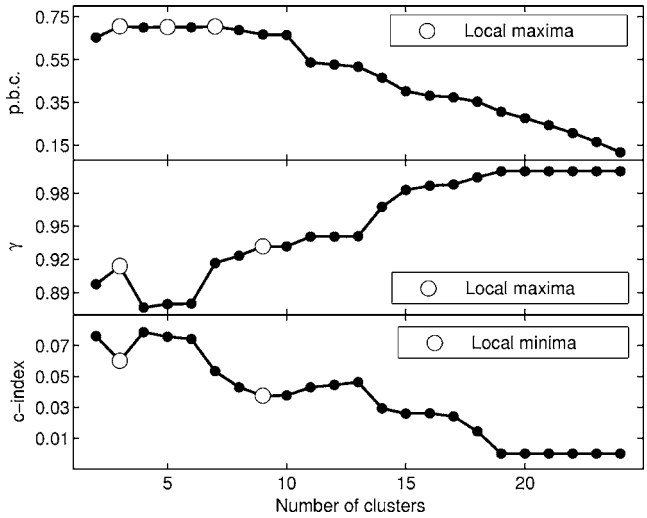


FIG. 4. Statistical indices used to evaluate the number of clusters present in the dataset across partitioning levels.

TABLE II. Contingency table for the modal response in the first cluster of participants ($N=21$). Material: S=Steel, G=Glass, W=Wood, P=plexiglass; Area: A1–A5=75–1200 cm^2. Response categories in italics.

	Metal					*Glass*				
S	1	3	17	21	21	20	18	4	0	0
G	2	2	18	21	21	19	18	3	0	0
W	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0
	Wood					*Plastic*				
S	0	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	0	0	0
W	19	13	10	13	2	2	8	11	8	19
P	21	21	16	6	3	0	0	5	15	18
	A1	A2	A3	A4	A5	A1	A2	A3	A4	A5

dices. The three clusters contained 21, 3, and 1 participant(s); data for these groups are shown in Tables II, III, and IV, respectively. For the first group, performance was almost perfect with respect to the gross material categories wood-plexiglass and steel-glass, with only one wood response being given for a glass sound. Also a strong tendency to associate the glass and wood responses with smaller plates and the metal and plastic responses with larger plates was found. The same tendencies characterized the second group of participants, the only difference being the association of the wood and plastic responses with larger and smaller wood or plexiglass plates, respectively. The participant in the third cluster consistently associated metal with large wood and plexiglass plates and did not have plastic among the modal responses. Subsequent statistical modeling was performed on data from the main group of participants.

The relevance of properties of the sound source to experimental judgment was first tested. Separate logistic regression models (Agresti, 1996) were built for the perceptual categorization within each of the gross material categories (metal/glass and wood/plastic responses). The wood modal response observed for one of the glass plates was not included in the analysis. Parsimonious models were sought, following the approach suggested by Hosmer and Lemeshow (1989). Thus, before entering predictors into multivariate models, the significance of their effect was tested within univariate models. Both the material, and the area of the plates, were coded as categorical variables. The models' goodness-of-fit was evaluated with the deviance and Hosmer-Lemeshow (Hosmer and Lemeshow, 1989) statistics, nonsignificant values indicating the statistical equivalence of observed and predicted data, and thus the validity of model-based inferences. On the other hand, given the almost perfect performance level observed, data concerning identification of the gross material categories could not be modeled using logistic regression. Simple χ^2 association tests were therefore adopted.

Identification of the gross material categories was influenced by the material, but not by the area of the plates [$\chi^2(3)=416.038$, $p<0.001$, $\chi^2(4)=0.038$, $p=1.000$, respectively]. Also, steel and glass plates were identified equally often as being made of metal or glass [$\chi^2(1)=1.005$, $p=0.316$] and wood and plexiglass plates were identified equally often as being made of wood or plastic [$\chi^2(1)=0$, $p=1$]. On the contrary, plate material did not influence significantly the identification within the gross categories [Wald $\chi^2(1)=0.052$, $p=0.820$, Wald $\chi^2(1)=1.962$, $p=0.161$, for metal/glass and wood/plastic, respectively], while the effect of area was highly significant in both cases [Wald $\chi^2(4)=47.386$, $p<0.001$; Wald $\chi^2(4)=48.52$, $p<0.001$, respectively]. Finally, the effect of area alone accounted well for the within-gross category identification data [metal/glass:

TABLE III. Contingency table for the modal response in the second cluster of participants ($N=3$). Material: S=Steel, G=Glass, W=Wood, P=plexiglass; Area: A1–A5=75–1200 cm^2. Response categories in italics.

	Metal					*Glass*				
S	0	3	3	3	3	3	0	0	0	0
G	0	1	2	3	3	3	2	1	0	0
W	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0	0	0
	Wood					*Plastic*				
S	0	0	0	0	0	0	0	0	0	0
G	0	1	0	0	0	0	0	0	0	0
W	0	1	2	3	2	3	2	1	0	1
P	0	0	0	2	2	3	3	3	1	1
	A1	A2	A3	A4	A5	A1	A2	A3	A4	A5

TABLE IV. Modal response for the participant in the third cluster. Material: S=Steel, G=Glass, W=Wood, P=plexiglass; Area: A1–A5=75–1200 cm^2. Response categories in italics (M=metal).

S	*G*	*M*	*M*	*M*	*M*
G	*G*	*W*	*M*	*M*	*M*
W	*W*	*W*	*W*	*M*	*M*
P	*W*	*W*	*W*	*M*	*M*
	A1	A2	A3	A4	A5

deviance=0, $p=1$, Hosmer-Lemeshow $\chi^2(2)=0.009$, $p=0.996$; wood/plastic: deviance=0, $p=1$, Hosmer-Lemeshow $\chi^2(3)=0$, $p=1$].

Discussion: Consistent with previous studies (Gaver, 1988; Kunkler-Peck and Turvey, 2000; Giordano, 2003; Tucker and Brown, 2003), nearly all individual listeners (88%) showed perfect identification of gross material categories (steel-glass and wood-plexiglass), independently of the geometry of the plates. From the mechanical point of view different material properties could explain this performance, steel and glass being both denser and stiffer than wood and plexiglass (see Table I and Waterman and Ashby, 1997).

Highly impaired performance was observed for the identification of materials within the gross categories: steel was perceptually equivalent to glass, wood to plexiglass. These results are consistent with those of Lutfi (2001), Giordano (2003), and Tucker and Brown (2003), but not with the perfect wood/plexiglass identification reported by Kunkler-Peck and Turvey (2000). Inconsistently with data from Kunkler-Peck and Turvey (2000) and Tucker and Brown (2003), identification within the gross categories was influenced by the geometry of the plates, glass, and wood being associated with smaller plates than metal and plastic. The possible sources for these inconsistencies are addressed in Sec. IV C.

In the Introduction, the effect of plate size on identification reported by Giordano (2003) was hypothesized to be due to the absence of acoustical differences between materials in the same gross category. It must however be pointed out that even in the absence of acoustical support for perfect recognition the observed strong concordance among listeners in associating material type with size points toward a cognitive origin for these effects, rather than to a procedural bias. Indeed, if these associations resulted from the tendency to focus on the only source property that carried significant acoustical variations (allegedly size), an equal number of participants would have associated given material types to opposite sizes. This was not the case in the current data.

B. From acoustics to perception

The acoustical basis for the perceptual categorization of materials within the same gross category was investigated, using the procedure outlined in Sec. IV A. A different approach was used for the identification of the gross categories and is presented in section IV C.

With regression models, the transform of the predictor affects its association with the predicted response. For each acoustical predictor a transform was chosen among the linear (identity transform), logarithmic and, for F, the ERB-rate transform, taking the absolute value of the slope-measures to evaluate the logarithmic transform. The univariate models with the different transforms of the same predictor were then compared on the basis of their log-likelihood. Thus, the chosen transform was that yielding the univariate model with the highest log-likelihood, i.e., the model closest to the unknown true probability distribution from which observations were sampled (cf. Golden, 2000). The results of this analysis are shown in Table V. The correlations among the acoustical indices transformed accordingly are shown in Table VI.

Given the presence of strong correlations among predictors, different regression models may, in principle, account for the same data. For this reason, model selection procedures that produce one single model in their output were not adopted (e.g., backward elimination, forward selection). It was thus decided to compute all possible models, starting from the univariate ones and progressively increase the number of predictors until at least one of the models was associated with a nonsignificant goodness-of-fit statistic.

For the metal/glass data set, F alone was sufficient to account for the observed data [deviance(8)=7.723, $p=0.259$; Hosmer-Lemeshow $\chi^2(8)=9.700$, $p=0.138$]. The probability of choosing the metal category increased with

TABLE V. Log-likelihood (LL) of the models computed to select the transform for the acoustical predictors. MG=metal/glass dataset; WP=wood/plastic dataset. The LL of the models with the selected transform is shown in boldface.

Data set	Acoustical descriptor	LL linear model	LL logarithmic model	LL ERB model
MG	$\tan\phi_{aud}$	−135.71	**−133.10**	
	Dur	−124.33	**−122.66**	
	F	−57.50	**−48.53**	−50.25
	Lou_{att}	−123.61	**−122.02**	
	Lou_{mea}	−86.70	**−85.29**	
	Lou_{sl1}	−89.08	**−88.37**	
	Lou_{sl2}	**−124.79**	−131.52	
	SCG_{att}	−87.93	**−87.84**	
	SCG_{mea}	−127.49	**−127.24**	
	SCG_{slo}	−119.08	**−108.97**	
WP	$\tan\phi_{aud}$	**−135.23**	−135.55	
	Dur	−126.59	**−121.59**	
	F	−105.30	**−101.79**	−103.34
	Lou_{att}	**−127.87**	−127.97	
	Lou_{mea}	**−130.25**	−130.85	
	Lou_{sl1}	−103.58	**−103.02**	
	Lou_{sl2}	**−134.52**	−138.72	
	SCG_{att}	−110.08	**−109.87**	
	SCG_{mea}	**−103.86**	−104.03	
	SCG_{slo}	**−118.63**	−121.96	

TABLE VI. Correlation among acoustical predictors, transformed according to the analysis summarized in Table V. The upper triangular matrix reports correlations for the wood/plastic data set, and the lower triangular matrix reports correlations for the metal/glass data set. Significant correlations (df=8, $p \leq 0.05$) are shown in boldface.

	WP									
MG	$\tan\phi_{aud}$	Dur	F	Lou_{att}	Lou_{mea}	Lou_{sl1}	Lou_{sl2}	SCG_{att}	SCG_{mea}	SCG_{slo}
$\tan\phi_{aud}$		**−0.692**	0.219	−0.49	0.499	0.099	−0.603	−0.047	0.178	−0.476
Dur	**−0.637**		−0.461	0.086	**−0.752**	−0.41	**0.844**	−0.377	**−0.649**	**0.718**
F	0.315	−0.471		**0.722**	0.172	**0.968**	−0.092	**0.917**	**0.802**	**−0.783**
Lou_{att}	**0.693**	**−0.783**	0.596		−0.142	**0.761**	0.345	**0.857**	0.616	−0.355
Lou_{mea}	**−0.804**	0.416	**−0.694**	−0.619		0.074	**−0.902**	0.164	0.493	−0.239
Lou_{sl1}	**0.81**	**−0.669**	**0.786**	**0.847**	**−0.929**		−0.026	**0.884**	**0.784**	**−0.732**
Lou_{sl2}	**0.8**	−0.191	0.456	0.455	**−0.934**	**0.778**		−0.021	−0.357	0.27
SCG_{att}	−0.265	−0.125	**0.749**	0.203	−0.239	0.322	−0.038		**0.852**	**−0.745**
SCG_{mea}	−0.534	−0.081	0.487	0.022	0.213	−0.036	−0.45	**0.764**		**−0.784**
SCG_{slo}	**0.792**	**−0.74**	0.455	**0.637**	**−0.685**	**0.744**	0.54	0.065	−0.354	

decreasing F. For the wood/plastic dataset, none of the acoustical predictors alone could account sufficiently well for observed responses. Five of the two-predictor models were, instead, associated with nonsignificant goodness-of-fit statistics [deviance(8) $\leq$ 13.533, $p \geq 0.060$; Hosmer-Lemeshow $\chi^2(8) \leq 11.948$, $p \geq 0.154$]. For the first two models, the most important predictor, i.e., that associated with the highest standardized parameter estimate, was F, the second predictor being either Lou_{mea} or Lou_{sl2}, whereas for the other three models, the most important predictor was Lou_{sl1}, the least important predictor being Lou_{mea}, Lou_{sl2}, or Dur. The probability of choosing the wood category increased with increasing F and Lou_{mea} and with decreasing Dur, Lou_{sl1}, and Lou_{sl2} (i.e., with faster loudness decays). It is worth noting that the primary parameters F and Lou_{sl1} are highly correlated for the wood/plastic dataset. Figure 5 shows the selected regression model for the metal/glass dataset and the F-Lou_{sl2} model for the wood/plastic dataset.

Discussion: Consistent with the results of Klatzky *et al.* (2000), the metal/glass identification was based on signal frequency, glass being associated with higher frequencies than metal. As pointed out in the Introduction, the impaired performance in the identification of hard materials reported by Lutfi and Oh (1997) was due to an excessive weighting of signal frequency. A similar explanation for impaired performance might apply here. The relevance of frequency for this categorization is, however, not consistent with results of Roussarie (1999). The simplest explanation for this inconsistency that emerges from a review of previous studies is based on the range of variation of frequency within the stimulus sets; seven semitones in Roussarie (1999), 4.05 octaves in the present study.

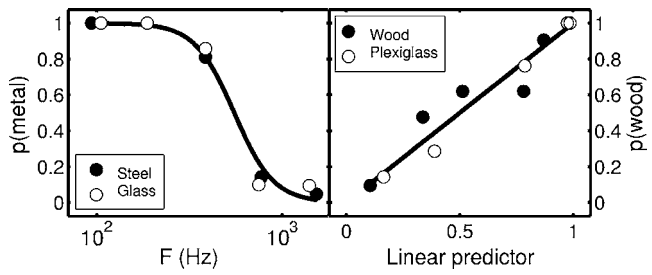


FIG. 5. Left panel: observed and predicted proportions of choosing the response *metal* as a function of the F parameter. Right panel: observed and predicted proportion of choosing the response *wood* as a function of the linear predictor in the F-Lou_{sl2} model.

Frequency was also found to explain the wood/plastic identification where, consistently with Klatzky *et al.* (2000), wood was associated with higher frequencies than plastic. However, this variable alone was not sufficient to account for the observed data, other necessary though secondary variables being either average signal loudness or Lou_{sl2}. The same data were also explained in terms of Lou_{sl1}, strongly correlated with F, and either duration, average loudness, or Lou_{sl2} as secondary variables.

Following a principle of parsimony, a common acoustical explanation for both the metal/glass and wood/plastic data was sought. It can be concluded, then, that both relied mainly on frequency, an acoustical parameter that also explains the relevance of plate size to the judgments. Frequency effects on dissimilarity rating were also reported by McAdams *et al.* (2004), although, as mentioned previously, Roussarie (1999) found material identification to be completely independent of frequency.

Finally, these analyses demonstrate that the investigated acoustical measure of damping, $\tan\phi_{aud}$, does not account for several auditory material categorizations.

C. From acoustics to physics

The presence of sufficient acoustical information for perfect material identification was ascertained. Given the almost perfect performance observed for the identification of the gross material categories, this analysis is equivalent to pointing out the possible acoustical criteria for judgment.

$\tan\phi_{aud}$ was already found to discriminate perfectly among all materials (see Sec. III). Concluding as to the presence of sufficient information on the basis of this result would be incautious if not incorrect, not the least because its perceptual relevance is questioned in the current study. Therefore this descriptor was not taken into account in the following analyses.

Logistic regression was used to find which acoustical descriptor or combination of descriptors allowed for perfect

TABLE VII. Acoustical descriptors found to categorize perfectly the contrasted materials. For each acoustical descriptor the sign of the association with the boldfaced category is also shown (e.g., the model in the bottom row shows that wood is associated with a combination of higher Lou_{att} and higher SCG_{slo} values).

Dataset	Acoustical descriptors
Steel-Glass	Dur(+)
vs	Lou_{att}(+)
Wood-Plexiglass	SCG_{slo}(+)
Steel	Dur(+)SCG_{att}(+)
vs	Dur(+)SCG_{mea}(+)
Glass	Dur(+)SCG_{slo}(+)
	F(+)Lou_{att}(−)
	F(+)Lou_{sl1}(+)
	F(+)SCG_{slo}(+)
	Lou_{att}(−)SCG_{att}(+)
	Lou_{att}(−)SCG_{mea}(+)
	Lou_{sl1}(+)SCG_{att}(+)
	SCG_{att}(+)SCG_{slo}(+)
	SCG_{mea}(+)SCG_{slo}(+)
Wood	F(−)Lou_{att}(+)
vs	Dur(+)Lou_{att}(+)
Plexiglass	Dur(+)Lou_{mea}(+)
	Dur(+)Lou_{sl1}(−)
	Dur(+)SCG_{mea}(+)
	Lou_{att}(+)Lou_{sl1}(+)
	Lou_{att}(+)SCG_{att}(−)
	Lou_{att}(+)SCG_{slo}(+)

material identification. In analogy with the analysis of behavioral data, separate analyses were performed for the identification of the gross material categories and for identification within the same gross category. Thus, using the descriptors transformed as in Sec. IV B, regression models were sought that resulted in so-called complete separation (i.e., perfect prediction of the dependent variable, material type; Albert and Anderson, 1984). Models were selected starting from the univariate cases and the number of descriptors was progressively increased until at least one perfectly identifying model was found. For each of the considered data sets, the final models produced a threshold above and below which materials belonged to one and only one category. This threshold was defined by a value of the acoustical descriptor in the univariate case or, for models including two descriptors, by a line in the plane defined by the acoustical parameters (see Fig. 5). Table VII reports the results of this analysis.

Steel-glass sounds were thus characterized by higher values of Lou_{att}, Dur, and SCG_{slo} than those characterizing the wood-plastic sounds. It is highly likely that at least one of these acoustical parameters, eventually including $\tan\phi_{aud}$, was used by participants for identification of the gross categories. Several pairs of descriptors perfectly categorized steel and glass (11 pairs) on the one hand, and wood and plexiglass (8 pairs) on the other. Overall, two pairs of descriptors perfectly identified all material types: Dur-SCG_{mea} and Lou_{att}-F. Figure 6 shows the optimal identification criteria based on the Dur-SCG_{mea} parameters. We are thus justified in concluding on the presence of sufficient acoustical information for perfect material identification.

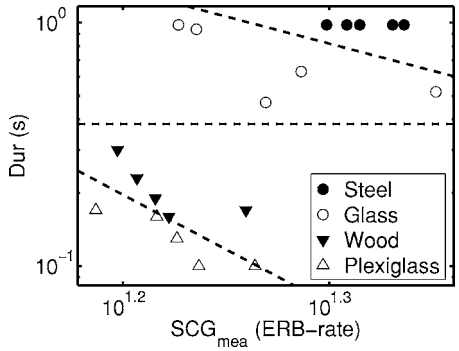


FIG. 6. Optimal criteria for material categorization. Dashed lines show the equal probability boundaries (thresholds) for the optimal criteria.

Finally, it is interesting to compare the optimal use of acoustical information with the observed behavioral criteria, focusing, in particular, on F. While optimal criteria associated steel and plexiglass with higher frequencies, participants used exactly the opposite weighting.

Discussion: Several acoustical parameters accounted for the perceptual (and optimal) identification of the gross steel-glass and wood-plexiglass material categories. Consistently with results by Klatzky *et al.* (2000) and Avanzini and Rocchesso (2001), $\tan\phi_{aud}$ had a lower value for the steel-glass signals. The same discrimination was also explained by Dur, Lou_{att}, and SCG_{slo}. Interestingly, the last of these parameters explained the perceptual relevance of damping in two of the experiments reported by McAdams *et al.* (2004). Given the presence of multiple acoustical explanations, no conclusions can be drawn on which of these parameters was actually used by listeners. In particular, it cannot be excluded that $\tan\phi_{aud}$ was attended to by listeners. Concerning within-gross category identification, the absence of perfect auditory performance is in contrast with the ability of $\tan\phi_{aud}$ to separate the different materials perfectly. It is thus evident that $\tan\phi_{aud}$ was not used by participants for these categorizations. Excluding $\tan\phi_{aud}$, several pairs of other acoustical descriptors allowed perfect within-gross category identification. Among them three were based on the main acoustical parameter used for perceptual categorizations: frequency. The optimal weighting of this parameter was, however, contrary to the observed perceptual weighting. Thus two causes for impaired performance can be hypothesized: the wrong weighting of signal frequency and the absence of focus on the other acoustical parameters necessary for perfect identification, such as attack loudness. Discriminating between these two alternatives is not possible with the available data.

The reasons for the inconsistency between the impaired wood/plastic identification reported in this study and by Giordano (2003) and Tucker and Brown (2003) and the perfect performance reported by Kunkler-Peck and Turvey (2000) still remain unclear. It should be noted however that Kunkler-Peck and Turvey (2000) did not use recorded signals, but generated them live, increasing the likelihood of making additional information for the material type available. For example, as the repetitions provided to the partici-

pants were not acoustically identical, extraction of material-specific invariant acoustical information might have been favored. Also, the manipulation of the plates necessary to hang them on the device after each trial might have generated additional acoustical signals (e.g., scraping sounds) that were potentially informative with respect to the object's material. Another inconsistency with previous studies concerns the effect of plate geometry on performance: it was secondary, at best, in Kunkler-Peck and Turvey (2000) or Tucker and Brown's (2003) data but was strong in the current data. Plausibly, the geometrical variation of the sources in these former studies came with less acoustical variation than in the present study, thus facilitating a reduced focus on this source property. Consistently, Giordano (2003) found plate area and not shape to be perceptually relevant, the former most likely causing stronger acoustical variations than the latter.

In summary, acoustical support for potentially perfect performance was highlighted. Thus, the relevance of plate area to material identification, also reported previously by Giordano (2003), is not a product of the absence of sufficient information for the task. Consequently the observed biases are likely to have a cognitive and not a methodological origin. Observed response profiles are indeed likely to reflect the regularities of the everyday acoustical environment (cf. Barlow, 2001). Concerning identification of gross material categories, available measures of the mechanical properties of engineering materials report plastics (polymers) and woods as strongly different from metals and glasses (Waterman and Ashby, 1997). Given these differences, it is highly likely that, independently of their geometry, signals originating from wood and plastic objects would always be differentiated from those originating from metal and glass objects. Consequently, the everyday perceiver would have a rather easy time learning to make robust identifications independent of object geometry, as found in this study.

The ecological explanation of the biases within gross categories is less clear. The simplest hypothesis is based on geometry. For example, the glass impact sounds experienced everyday are probably generated by smaller objects than is the case with metal objects (e.g., klinking glasses vs banging pans) and large, freely vibrating glass objects, such as the plates of the current study, would be too fragile to be of any plausible ordinary use. One might object that sounds generated by striking small metallic objects (coins or keys) are also frequently experienced. These signals, however, have a more complex nature than those investigated in the current study, comprising multiple rather than single impacts, eventually interleaved with signals generated by nonimpact interactions among objects (e.g., friction). Assuming as illegitimate the generalization of source recognition criteria from one kind of object interaction to the other, the size explanation for the metal/glass perceptual identification still seems valid. However, it does not appear convincing for the wood/plastic identification.

Given the plausible high relevance of frequency to the wood/plastic categorization, any source property that significantly affects this signal property might be a potential candidate to explain the size bias. Increasing modal frequencies come, for example, with a decrease in size and density and with an increase in thickness and Young's modulus or simply in stiffness (cf. Fletcher and Rossing, 1991). Therefore it might be hypothesized that listeners learn to associate wood with higher frequencies than plastic because the wood sounds we experience every day are generated by thicker objects than for plastic sounds. This hypothesis appears plausible, given that thin layers should be more easily manufactured with plastics than woods, but could be hardly generalized to the metal/glass case. Concerning Young's modulus, one should expect woods and glasses to be stiffer than, respectively, plastics and metals. Such differences, however, are not apparent in published measures of engineering materials (Waterman and Ashby, 1997). Concerning density, glasses and woods would then be expected to be less dense than metals and plastics, respectively. Indeed, with published measures, the average densities of these materials follow this order (Waterman and Ashby, 1997). The use of an identical explanation for both the metal/glass and wood/plastic categorizations makes this hypothesis particularly attractive.

V. CONCLUSIONS

Material identification from impact sounds was investigated. All the pairwise relationships between source, signal and recognized source properties were studied.

Analysis of the relations between source properties and recognition performance highlighted perfect identification of the gross material categories steel-glass and wood-plexiglass. However, impaired categorization of materials within the same gross category was observed, material identification relying only on the size of the objects. A strong agreement between individuals was also observed.

Acoustical criteria for material identification were investigated. Previous studies found identification to be influenced by acoustical measures of damping. Therefore, a psychoacoustically inspired measure of damping, $\tan\phi_{\text{aud}}$, was contrasted with a large set of signal descriptors in its ability to explain the behavioral data. This measure was found to account only for the identification of the gross material categories, the same data being equally well accounted for by other signal properties: duration, attack loudness, and decay rate of the spectral center of gravity. Identification within the gross categories was instead found to be based mainly on signal frequency, although the wood/plastic identification was equally well accounted for by loudness decay descriptors and, as secondary variables, signal duration or average loudness. Thus, only partial support for the perceptual relevance of $\tan\phi_{\text{aud}}$ was found.

Analysis of the relationship between acoustical and source properties highlighted the presence of sufficient information for perfect material identification, pointing toward the cognitive origin of the observed biases. Identification data were thus interpreted with reference to the regularities of the everyday acoustical environment.

ACKNOWLEDGMENTS

Bruno L. Giordano was supported by the European Commission's Future and Emergent Technologies collaborative R&D program, Project No. IST-2000 25287 (Sob-

The Sounding Object), and by a Marie Curie Training Site Fellowship hosted by KTH Music Acoustics in Stockholm, FP5 Contract No. HPMT-CT-2000-00119. This project was supported in part by a grant from the French Ministry of Research's Cognition and Information Processing Program to Stephen McAdams. The authors wish to thank Alain de Cheveigné for providing the code implementing the simulation of the peripheral auditory system processing and for suggestions concerning the $\tan\phi_{\text{aud}}$ extraction procedure.

Agresti, A. (**1996**). *An Introduction to Categorical Data Analysis* (Wiley, New York).

Albert, A., and Anderson, J. A. (**1984**). "On the existence of maximum likelihood estimates in logistic regression models," Biometrika **71**(1), 1–10.

Avanzini, F., and Rocchesso, D. (**2001**). "Controlling material properties in physical models of sounding objects," *Proceedings of the International Computer Music Conference 2001*, La Habana, Cuba, pp. 91–94.

Baker, F. B., and Hubert, L. J. (**1972**). "Measuring the power of hierarchical cluster analysis," J. Am. Stat. Assoc. **70**, 31–38.

Barlow, H. (**2001**). "The exploitation of regularities in the environment by the brain," Behav. Brain Sci. **24**, 602–607.

Carello, C., Wagman, J. B., and Turvey, M. T. (**2003**). "Acoustical specification of object properties," in *Moving Image Theory: Ecological considerations*, edited by J. Anderson and B. Anderson (Southern Illinois University Press, Carbondale).

Fletcher, N. H., and Rossing, T. D. (**1991**). *The Physics of Musical Instruments* (Springer-Verlag, New York).

Gaver, W. W. (1988). "Everyday listening and auditory icons," Unpublished doctoral dissertation, University of California, San Diego.

Giordano, B. L. (**2003**). "Material categorization and hardness scaling in real and synthetic impact sounds," in *The Sounding Object*, edited by D. Rocchesso and F. Fontana (Mondo Estremo, Firenze), pp. 73–93.

Golden, R. M. (**2000**). "Statistical tests for comparing possibly misspecified and nonnested models," J. Math. Psychol. **44**, 153–170.

Gordon, A. D. (**1999**). *Classification* (Chapman and Hall/CRC, Cleveland).

Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**(5), 1270–1277.

Guski, R. (2000). "Studies in auditive kinetics," in *Contributions to Psychological Acoustics. Results of the 8th Oldenburg Symposium on Psychological Acoustics*, edited by M. M. A. Schick and C. Reckhardt BIS, pp. 383–401.

Hartmann, W. M. (**1997**). *Signals, Sound and Sensation* (AIP Press, Woodbury, NY).

Hosmer, D., and Lemeshow, S. (**1989**). *Applied Logistic Regression* (Wiley, New York).

Hubert, L. J., and Levin, J. R. (**1976**). "A general statistical framework for assessing categorical clustering in free recall," Psychol. Bull. **83**, 1072–1080.

Killion, M. C. (**1978**). "Revised estimate of minimum audible pressure: Where is the 'missing 6 dB'?," J. Acoust. Soc. Am. **63**(5), 1501–1508.

Klatzky, R. L., Pai, D. K., and Krotkov, E. P. (**2000**). "Perception of material from contact sounds," Presence: Teleoperators and Virtual Environment **9**(4), 399–410.

Kunkler-Peck, A. J., and Turvey, M. T. (**2000**). "Hearing shape," J. Exp. Psychol. Hum. Percept. Perform. **26**(1), 279–294.

Lambourg, C., Chaigne, A. and Matignon, D. (**2001**). "Time-domain simulation of damped impacted plates. II. Numerical model and results," **109**, 1433–1447,

Li, X., Logan, R. J., and Pastore, R. E. (**1991**). "Perception of acoustic source characteristics: Walking sounds," J. Acoust. Soc. Am. **90**(6), 3036–3049.

Lutfi, R. A. (**2001**). "Auditory detection of hollowness," J. Acoust. Soc. Am. **110**(2), 1010–1019.

Lutfi, R. A., and Oh, E. L. (**1997**). "Auditory discrimination of material changes in a struck-clamped bar," J. Acoust. Soc. Am. **102**(6), 3647–3656.

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**(5), 2946–2957.

McAdams, S. (**1993**). "Recognition of sound sources and events," in *Thinking in Sound: The Cognitive Psychology of Human Audition*, edited by S. McAdams and E. Bigand (Oxford University Press, Oxford), pp. 146–198.

McAdams, S. (**2000**). "The psychomechanics of real and simulated sound sources," J. Acoust. Soc. Am. **107**(5), 2792.

McAdams, S., Chaigne, A., and Roussarie, V. (**2004**). "The psychomecanics of simulated sound sources: Material properties of impacted bars," J. Acoust. Soc. Am. **115**(3), 1306–1320.

Milligan, G. W. (**1980**). "An examination of the effect of six types of error perturbation on fifteen clustering algorithms," Psychometrika **45**, 325–342.

Milligan, G. W. (**1981**). "A Monte Carlo study in thirty internal criterion measures for cluster analysis," Psychometrika **46**(2), 187–199.

Milligan, G. W. (**1996**). "Clustering validation: Results and implications for applied analyses," in *Clustering and Classification* edited by P. Arabie, L. J. Hubert, and G. De Soete (World Scientific, River Edge), pp. 341–375.

Milligan, G. W., and Cooper, M. C. (**1985**). "An examination of procedures for determining the number of clusters in a data set," Psychometrika **50**(2), 159–179.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**(3), 750–753.

Patterson, R. D., Allerhand, M., and Giguère, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**(4), 1890–1894.

Plack, C. J., and Moore, B. C. J. (**1990**). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**(5), 2178–2187.

Roussarie, M. V. (1999). "Analyse perceptive de structures vibrantes simulees par modele physique," Unpublished doctoral dissertation, Université du Maine, France.

Tucker, S., and Brown, G. J. (2003). "Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds," in *Aes 2003*.

van den Doel, K., and Pai, D. K. (**1998**). "The sounds of physical shapes," Presence **7**(4), 382–395.

Vanderveer, N. J. (1979). "Ecological acoustics: Human perception of environmental sounds," Unpublished doctoral dissertation, Cornell University. [Dissertation Abstracts International 40, 4543B. (University Microfilms No. 80-04-002).].

Waterman, N. A., and Ashby, M. F. (**1997**). *The Materials Selector*, 2nd ed. (Chapman and Hall, London).

Wildes, R., and Richards, W. (**1988**). "Recovering material properties from sound," in *Natural computation*, edited by W. Richards (MIT press, Cambridge), pp. 356–363.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Springer-Verlag, New York).

A structurally guided method for the decomposition of expression in music performance

W. Luke Windsor
School of Music and Interdisciplinary Centre for Scientific Research in Music, University of Leeds, Leeds LS2 9JT United Kingdom

Peter Desain
NICI, Radboud University, Postbus 9104, 6500 HE Nijmegen, The Netherlands

Amandine Penel
Laboratoire de Psychologie CognitiveUniversite de Provence & CNRS UMR 6146 Bat 9, Case D 3, place Victor Hugo 13331 Marseille Cedex 3 France

Michiel Borkent
NICI, Radboud University, Postbus 9104, 6500 HE Nijmegen, The Netherlands

(Received 25 March 2005; revised 12 October 2005; accepted 8 November 2005)

A method for separating, profiling, and quantifying the contributions of different structural components to expressive musical performance is described. The method is demonstrated through its application to a set of expert piano performances of a short piece from the classical period. The results show that the output of the method aids in the understanding of how the different structural components in a piece of music combine in the generation of an expressive performance. A second demonstration applies the method to performances at different tempi to illustrate its effectiveness in pinpointing the structural features responsible for small but statistically significant differences between performances. The method is compared with other approaches to the analysis and modeling of musical performance, and a number of potential applications are identified. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146091]

PACS number(s): 43.75.Cd, 43.75.St, 43.75.Yy [DD] Pages: 1182–1193

I. INTRODUCTION

A. Expression and structure in musical performance

Much progress has been made in the development of methods for extracting and analyzing discrete and continuous expressive parameters from audio and MIDI recordings of musical performances, and such methods have been used to develop and test hypotheses regarding the cognitive and motor processes which underlie such performances. Since the research of Seashore and colleagues (Seashore, 1938), it has been understood that skilled performers manipulate expressive parameters in their performances in structured and predictable ways that are related to the structure of the music. The vast majority of researchers in this area have concluded that many aspects of expressive timing and dynamics can be predicted from an analysis of the structure of a piece of music to be performed, and that such predictions are concrete enough to be formalized in a system of rules (see, e.g., Clarke, 1988; Palmer, 1997).

B. Generative approaches to expression in performance

The idea that the expressive aspects of musical performance are created from a representation of musical structure has led a number of researchers to advance computational theories that formalize and express the mapping from score plus structure to performance in algorithmic terms. We call these computational models generative theories here. For example, Clynes (1983) predicts timing and dynamics from time signature and composer, recursively subdividing time intervals multiplicatively at each metrical level. Friberg (1991; also see Sundberg, 1988) focuses on local structure (e.g., a jump in pitch) to calculate expressive deviations from the mechanical rendition. This approach uses a wide range of rules, each instantiating a different aspect of expression, the rules' effects accumulating in ways that may be quite difficult to interpret. Todd (1985, 1992, 1995) predicts timing and dynamics from phrase structure alone, applying a single formula (a parabola) additively at each level, an approach which has a recursive elegance. Such generative theories seem to have a huge advantage over other, less precise, theories (Desain *et al.*, 1998). One of their major benefits is that they can be fitted to empirical data, yielding an estimate of their predictive power and optimal parameter values. Such comparisons have been fairly widespread in the literature (e.g., Todd, 1992; Friberg, 1995; Windsor and Clarke, 1997; Widmer and Tobudic, 2003; also see Sec. I C). In most cases an overall measure of goodness of fit, or conversely a measure of error, is used to quantify the success with which a model (and, one assumes the theory upon which it is based) can explain an individual performance or set of performances.

However, although such generative computational models have greatly helped in building and testing the theoretical concepts used in the field, and are sometimes quite satisfactory in terms of output simulation (as in Widmer and Tobudic, 2003), they are in general not very successful when fits

to real performance data are attempted. This can be caused by the fact that many models are only partial, and expressive deviations linked to ignored types of structure easily upset the fitting process. For example, a local mid-bar phrase ending that is expressed with a ritardando in a performance would easily upset the optimization of Clynes' rather subtle composer's pulse, which is linked to the metric structure alone. Moreover, as empirical findings have demonstrated, identical effects might derive from very different rules or structures. A rule which maps the score location of an event within a phrase to a local modification of tempo can produce an effect which is indistinguishable from a rule which makes a similar prediction on the basis of metrical location. Similarly, a pause at or near a phrase boundary might be the result of a rule which applies to only one event (e.g., a micropause) or might be the result of a rule which applies to more than one event (e.g., a ritardando) (see Windsor and Clarke, 1997).

Note that this is not a criticism of generative theories as such, which can and sometimes do combine many different assumptions about expression. Musical structure seems not to be made of singular and homogeneous aspects, but constitutes a bundle of interlinked properties, which are often incompatible but not independent of each other. This complexity and interdependency has to be taken into account in investigating how musical structure gives rise to the expressive signal. What this paper addresses is how to better examine and quantify these multiple contributions to expression.

C. Estimating combined and individual fit of structural parameters

One solution would be to consider many kinds of musical structure at once and fit them jointly to a performance. Not only does this solve the problem of confounding ignored factors in the fitting procedure, but it also becomes possible to assess the relative contributions of different types of musical structure for a single piece. This was proposed by Desain and Honing (1997), and the current paper is an elaboration, implementation, and test of those ideas.

Given that a single piece may be structurally ambiguous and performers may even apply different strategies in relation to the same structure (see, e.g., Clarke and Windsor, 2000), these aspects constitute the so-called "interpretation" chosen by the performer and they form a rather important aspect of the data. This solution has been adopted with some success (such as in Sundberg *et al.*, 2003; Zanon and de Poli, 2003a, b), usually with quite specific (and quite local) rules that contain elaborate domain knowledge (like generating a pause before a large melodic leap) but only few parameters per rule. Our approach is different in that we do not aim to test any such specific aspects of expression. Instead, we assume regularity (e.g., each bar is expressed by the same timing fluctuation) and an open shape with a number of parameters (piecewise linear profiles) and aims to analyze expression (in this case expressive timing) in order to reveal more global mappings between structure and expression.

The method proposed here, which is implemented in the POCO environment [a software environment for the analysis of expression, see Honing (1990, 1992)] in a module entitled DISSECT (with SECT standing for Structural Expression Component Theory), not only delivers the relative contribution of the various components to the overall expressive profile, but also yields the component profiles themselves as well, effectively decomposing expression into its structural elements (note that to run POCO requires Macintosh Common LISP; for plotting results the scriptable statistics package JMP is used). Such decomposition may help to better reveal processes underlying the relationship between structure and expression. For example, if one measures the interonset timing of a number of performances of the same piece obtained under different conditions or from multiple performers, and merely compares the data in terms of their global differences or similarities [using the kinds of statistical methods applied by Shaffer (1981) or Repp (1992)] one is left with a rather uninformative result in regard to the underlying processes. It could be that there are systematic differences between performances (1) that reflect a difference in the application of various rules (e.g., a performer not expressing the time signature by means of timing); (2) that reflect the application of the same rules with different parameter settings or weights (a performer slowing down more or less in a phrase final ritard); or (3) that reflect the operation of the same rule on a different structural interpretation (e.g., expressing a different phrase structure with the same ritards at the end of each phrase). With a technique to decompose expression and compare its elements it becomes possible to distinguish between these hypotheses in a quantitative manner. Together with a few other attempts to judge the relative contributions of different musical structures in a systematic analysis (such as Thompson and Cuddy, 1997; Penel and Drake, 1998; Chaffin and Imreh, 2002; Sundberg *et al.*, 2003; Zanon and de Poli, 2003a, b), this method is high dimensional. It can be opposed to the visualization techniques applied to performance expression as elaborated by, for example, Dixon *et al.* (2002), which aim to represent expressive variation in a single time-variant plot of a few attributes like overall tempo and loudness. Although most generative theories propose quite explicit forms or shapes that make up the expressive signal (parabolic beat intervals, micropauses, recursive metric subdivisions), DISSECT works without imposing an explicit set of *a priori* expressive rules, hence it can be seen as more data driven. It does assume that the mapping from score to performance is constrained within parameter consistency; in other words, our assumption is that if an element of musical structure maps score to performance in a particular way, this relationship will be preserved for all examples of that structure within a performance. Secondly, we have chosen to assume that tempo change is linear (although the approach is not restricted to linear mappings in principle or practice). Hence, although our method has similarities to that described by Zanon and De Poli (2003a, b), it differs in that their approach is specific to a particular rule-based model of expression, whereas our approach is more general in formulation in that it evaluates a structural analysis of a piece and a mapping between this analysis and the expression, making only few *a priori* assumptions about what form the mapping might take.

Although the focus here is on expressive timing, our

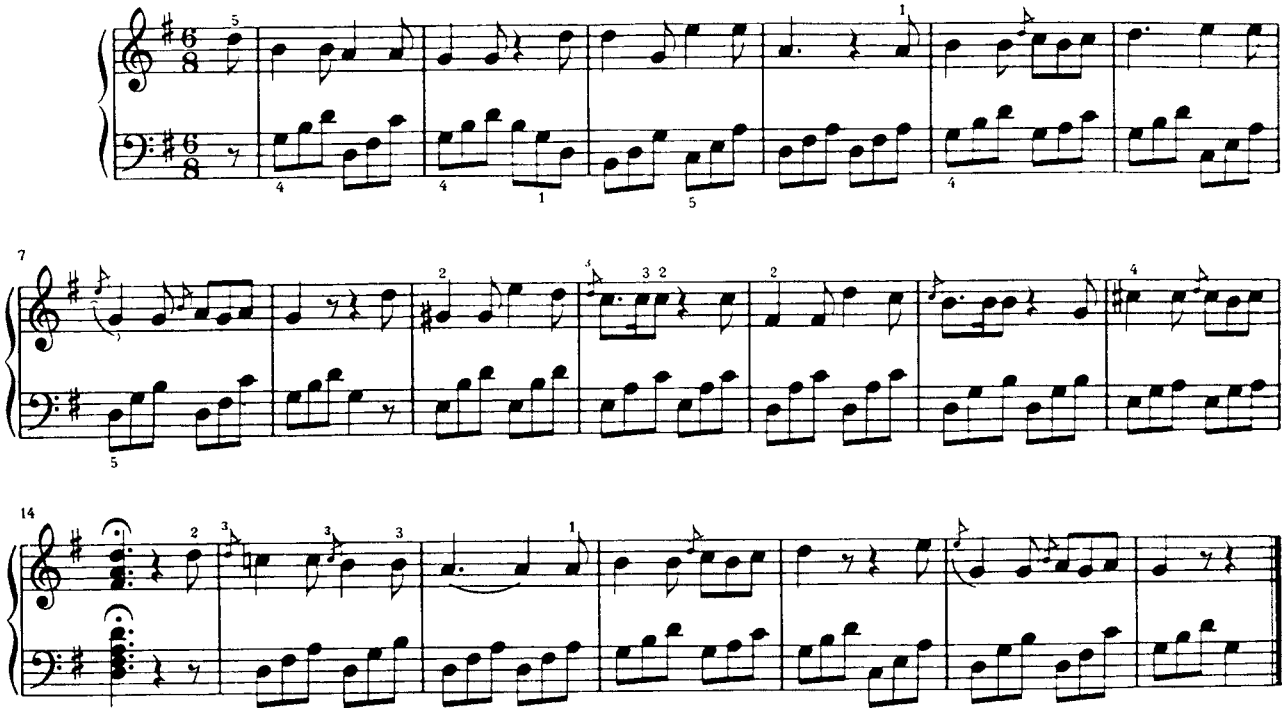


FIG. 1. Score of the Beethoven *Paisiello* theme.

method is in principle applicable to any expressive parameter, and this focus is chosen on pragmatic grounds. Moreover, although the dataset analyzed here was collected using a MIDI piano, the method can be applied to time series of measurements derived from an audio representation. The remainder of this paper demonstrates the application of DISSECT to a dataset of expert piano performance by analyzing the structural components contributing to performances at a single tempo, then showing how the deviations from proportional tempo which occur when a pianist is instructed to play at a higher or lower base tempo (see, e.g., Schmidt, 1985; Gentner, 1987; Desain and Honing, 1994; Repp, 1994; Windsor *et al.*, 2001) can be associated with differences in the interpretation of a small number or structural components.

II. THE TARGET DATASET OF PERFORMANCES

The performances modeled in this paper are derived from an earlier study which focused on grace note timing and the proportional tempo hypothesis (Windsor *et al.*, 2001) and are the same performances modeled in Penel *et al.* (1999) and Penel (2000). The piece performed has also been used in an earlier study of these issues (Desain and Honing, 1994).

A. Score

The piece used is the theme from Beethoven's six variations in G-major WoO 70 (1795) on the duet "Nel cor più non mi sento" from the opera "La Molinara" by Giovanni Paisiello (see Fig. 1).

The theme has a nominally isochronous broken-chord accompaniment in the left hand and a melody in the right, embellished by ornamental grace notes, and is notated in compound duple meter. The melodic gestures begin with an upbeat eighth note. The piece is essentially in two voices, except at the paused chord two-thirds through. Interestingly, the metrical and phrase structures of the piece are out of phase by one eighth-note unit, a common feature of music from this period. This feature alone suggests that this piece is an interesting candidate for the analysis to be carried out here, given that these two structural components might both be regarded as having a role to play in generating expression.

B. Performer and recording procedure

The performances were originally recorded for Windsor *et al.* (2001). The performer was a professional pianist and instrumental professor at the Tilburg Conservatory in the Netherlands (age 26). He was paid an appropriate hourly professional fee. The inter-onset timings of note onsets in the performances were captured using a Yamaha Disklavier MIDI grand piano and recorded via MIDI on a Macintosh PowerPC 9600/233 running a commercial sequencer package.

The performer had been given three weeks to prepare performances at nine different tempi from the score in Fig. 1. From these nine tempi, we have selected three instructed tempi for this study: "slow" [50 dotted quarter-note beats per minute (BPM)], "medium" (57 BPM), and "fast" (75 BPM). Although 50 BPM might seem rather slow for this piece, and 75 BPM rather fast, the pianist reported that they were musically acceptable. The "medium" tempo was regarded the most musically uncontroversial by the pianist.

Within the original experiment the pianist played randomized blocks of five repetitions of the theme at each of the tempi, giving a total of 45 complete performances. The pianist was allowed to practice the theme at the tempo requested

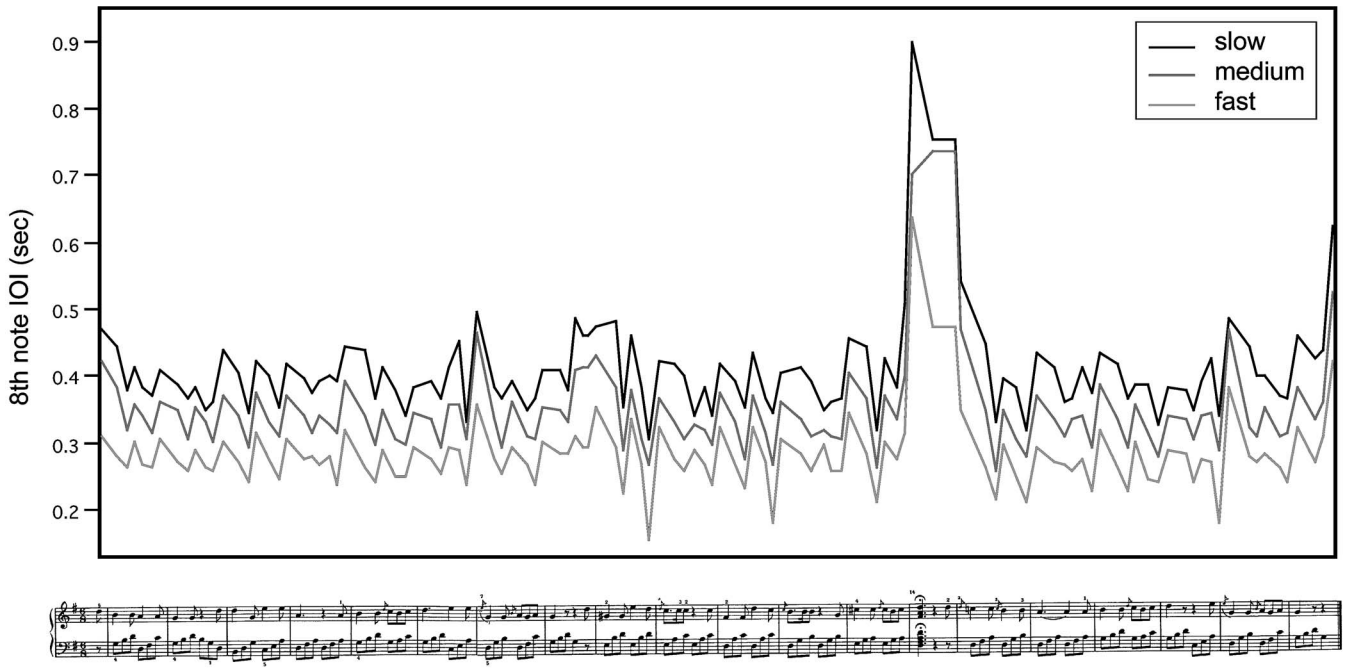


FIG. 2. Performances at each tempo, plotting score position against eighth-note IOI in seconds.

(a digital metronome was provided to remind the pianist of the tempo), and was asked to indicate whenever the next block could be recorded. Between each repetition there was a short break of about 5 s.

Using POCO (Honing, 1990) the onset times of all notes in the 15 performances were extracted, and inter-onset intervals (IOIs) were determined by onsets of melody notes (right hand) or by onsets of notes in the accompaniment (left hand) when there was no melody note. Grace note onsets were excluded from all analyses reported here [see Windsor *et al.* (2001) for an analysis of their timing].

C. Descriptive statistics for the selected performances

The 15 performances selected here were remarkably consistent within tempo condition, but show evidence of an effect of tempo on note timing. An ANOVA taking note IOI (for all onsets except those which precede rests in the score and the last onset) as the dependent variable, tempo condition and note position as factors, and repetition (five levels) as a random factor shows a significant interaction between note position and tempo condition ($F_{220,1320}=3.2153$, $p<0.0001$). Clearly, the performer did not maintain proportional timing over tempo at the note level, but was able to provide consistently timed performances within tempi. Hence, for the purposes of this paper average IOIs were calculated for each onset within each tempo, creating the three timing profiles shown in Fig. 2.

Comparison of the three profiles illustrates how well they correlate (about 0.95 between fast and medium and between medium and slow, and about 0.9 between fast and slow, $n=113$), despite having clear local differences for certain note positions and an offset due to the effect of global tempo.

III. APPLYING THE METHOD TO THE TARGET DATASET

A. Overview

The method, the statistical assumptions of which are outlined below, fits a generalized linear model to a time series of inter-onset intervals collected from a real performance. This model takes as its input a representation of the musical structures which might account for variation in expressive timing, estimates the fit, and provides prediction profiles for each element in this structure. The analysis can be thought of as a decomposition of the expressive timing into profiles associated with different kinds and levels of musical structure.

B. Assumptions and procedure

The method assumes that the expressive timing signal, expressed as beat length (inverse tempo), is a sum of several repeating (and possibly overlapping) timing profiles, each one reflecting the expression of a distinct structural unit such as a (sub)phrase or a metrical level. A subset of these units may form a hierarchical decomposition (e.g., bars and beats for a tight hierarchical structure), but this is not forced. The profiles are assumed to consist of line segments, with breakpoints specified at the first and last notes they span and, if necessary, at one or more intermediate notes (usually one extra breakpoint in the middle suffices).

Figure 3 shows a schematic depiction of a score, its structural annotation, and a profile for each structural unit. Note how each profile is determined by a set of breakpoints: the local tempo at each score time unit is estimated as a parameter in the fitting procedure to a real performance. In this sense the method is music-theoretically informed, because this structural description of the piece has to be pro-

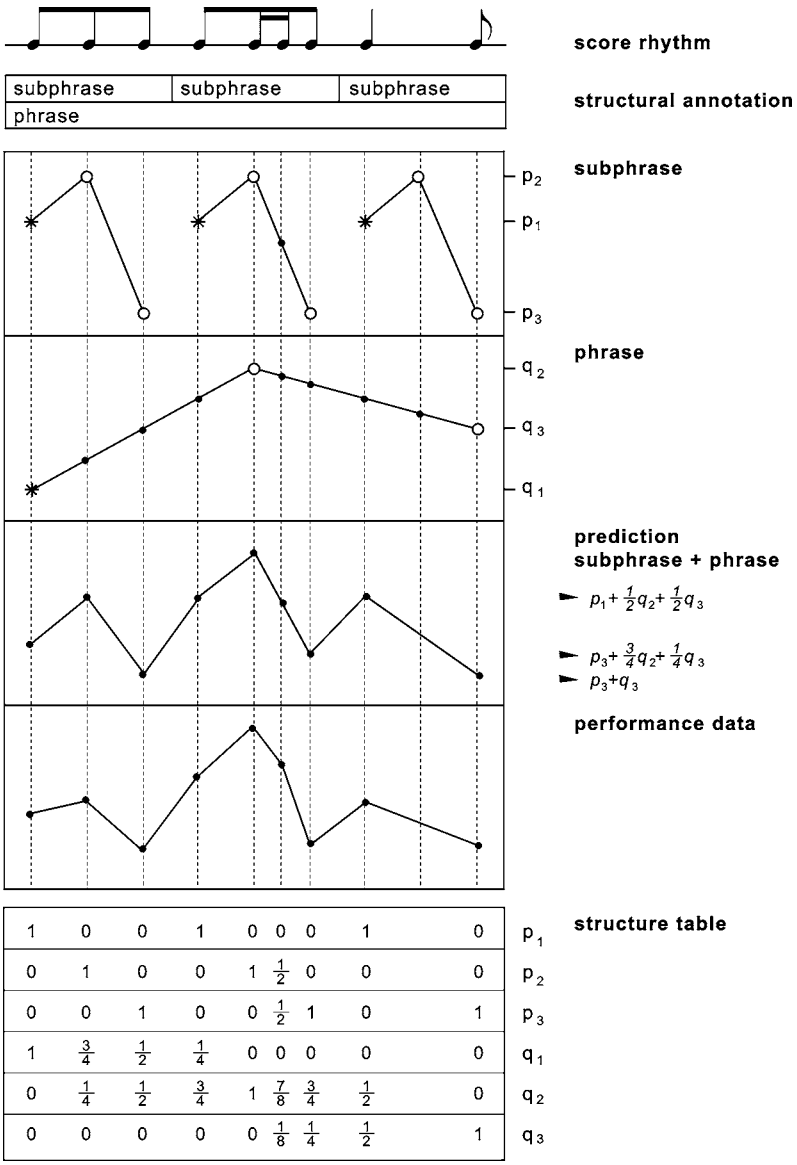


FIG. 3. A schematic depiction of a score, its structural annotation, the profiles for each structural unit, and how they combine into a prediction to be compared with performance data. The bottom panel illustrates how the structure is expressed as a matrix reflecting the linear combinations of parameters that constitute the model and is used for the regression analysis.

vided before an analysis can be done. The component profiles are defined by parameters, and the points of the overall profile are given by (weighted) sums of parameters, interpolating between them where necessary. The weights, which capture the structural description, are collected in a matrix A. There is a row in this table for each parameter, and a column for each note, i.e., each measured performance data point. The coefficients in the table specify the structural decomposition. If a note falls on a breakpoint of a profile, the corresponding coefficient in the table is 1; if it is outside the line segment starting or ending at that breakpoint, it is 0; and if it is on such a line segment, the coefficient expresses a linear combination (interpolation) of two parameter values. This table is generated from the structurally annotated MIDI score file in POCO. Now the predicted overall profile can be fit to the performance data.

If the expressive data to be predicted are expressed as vector x, with x_i being the local tempo of note i, and the parameters as vector p, the problem is to find the p_{opt} that minimizes the difference between the predicted A^*p and observed x:

$$p_{opt} = \text{argmin}_p \|A^* p - x\|.$$

Using the sum of the square errors as a measure of difference this is a linear regression problem that can be solved with simple means. The predicted overall profile is given by A^*p_{opt}. As the parameters p_i decompose into subsets, one set for each component, each component profile is calculated in a similar way, but zeroing in p_{opt} all parameters not belonging to that profile.

Since profiles repeat, we can usually create a nondegenerate matrix A and use fewer parameters than data points. However, because beginnings and ends of overlapping profiles will often coincide, the rank of A may be lower than its dimension. Clamping a few parameters to zero solves this problem.

The significance of individual parameters is not so relevant, as they form an inherent part of a profile, but the whole profiles are reanalyzed in a standard multiple regression which yields their contributions to the explained variance and their significance levels.

If optimization of free parameters leads to a good overall

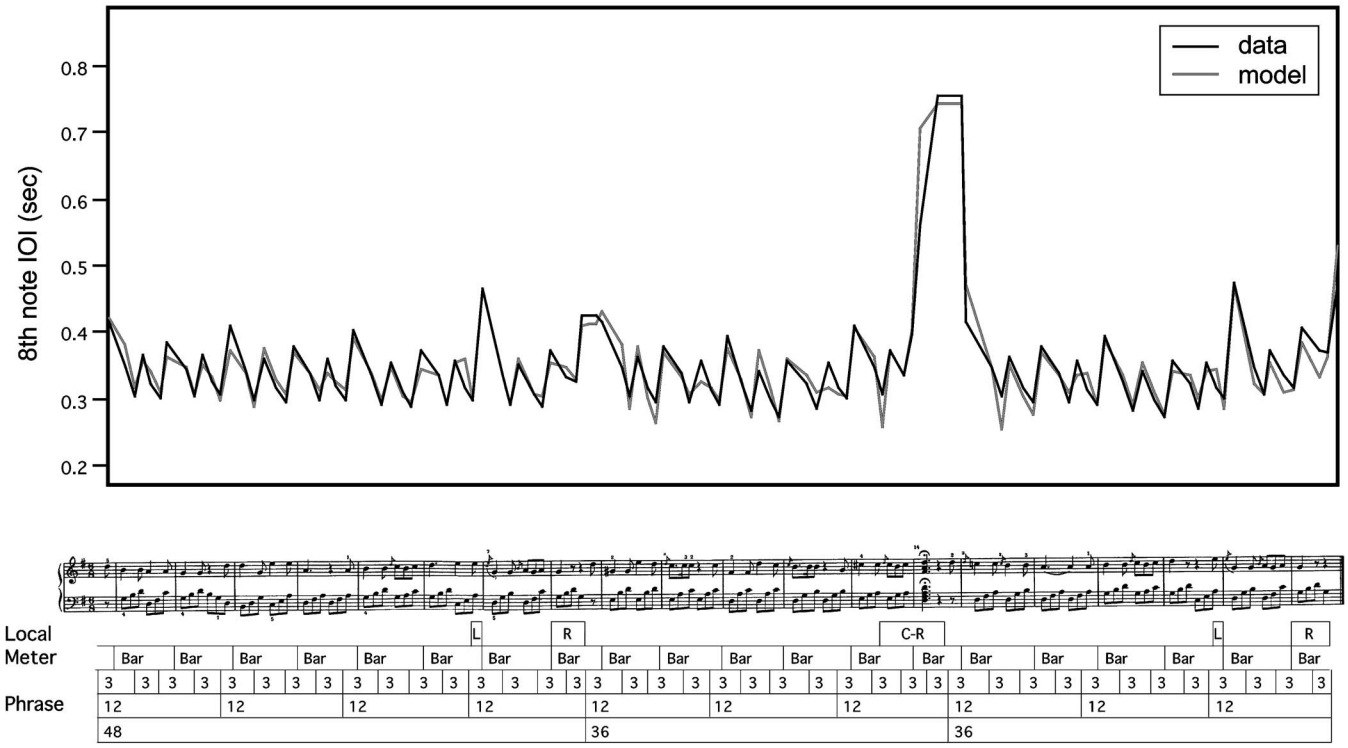


FIG. 4. Observed and predicted IOIs plotted against a score annotated with phrase, bar, and noncontiguous segments (L=leap; R=ritard; C-R= chord-ritard). Phrase segments are identified by their duration in score time measured in eighth notes. Note that the *x* axis is warped to align with the musical notation.

fit accounting for a large proportion of the variance, the musical structure is appropriately chosen. This means that the performance data exhibits systematic expressive features directly linked to the structural description.

C. The musical structure and constraints on the associated profiles

1. The structural representation

A set of structural units was added to the score in POCO (which adds a structural annotation capability to standard MIDI files). This structural annotation reflects the analytical intuitions of the first and third authors, breaking the piece down into a hierarchy of phrases, a metrical hierarchy, and identifying local sources for expression at phrase endings and accounting for the fermata. The analysis here is similar to that employed in Penel *et al.* (1999) and Penel (2000). The score is annotated with these structural units in Figs. 4 and 5, and further detail on the segmentation can be found in Table I.

2. Profile constraints

We have chosen here to constrain the model to a certain extent in order to reduce the number of free parameters in line with some hypotheses about patterns of expressive timing. The following constraints represent the "generative rules" we have chosen to include: (1) tempo change is linear between breakpoints (see below for our rationale for this) and (2) expressive timing is equal when structural units are repeated. In this instance each structural unit has a profile consisting of straight line segments with breakpoints at determined positions. The program allows for arbitrary complex shapes with many breakpoints, but only three point profiles (for the main phrase and metric units), two point profiles (for the final ritard), and local one point units (for local effects) were used, except for the profile for 12-phrase, which has five breakpoints to allow for expression associated with its initial upbeats and final interval. Table I specifies the extent, shape, and number of free parameters for each profile with brief descriptions. All other intermediate breakpoints are located at the midpoint of the structural unit. Other shapes or linear combinations could have been applied to the analysis of this performance, but small numbers of breakpoints and piecewise linear profiles were adopted to demonstrate the application a set of relatively simple statistical assumptions. Many generative models use nonlinear curves, but these tend to mask step tempo changes [see Windsor and Clarke (1997) for a discussion of this issue in relation to Todd (1992)].

3. Accounting for global tempo

The resulting profiles combine additively to predict the observed performance. However, the freedom in doing so is still too large: the global tempo can be explained as an offset to any profile that spans the whole piece, or be distributed between them. Hence, choices have to be made to reduce the number of parameters and make a unique solution to the optimization possible. In this instance the initial values of most profiles are clamped to zero (except for the global tempo intercept and the local pointwise parameters), and a constant intercept parameter was added to capture the global

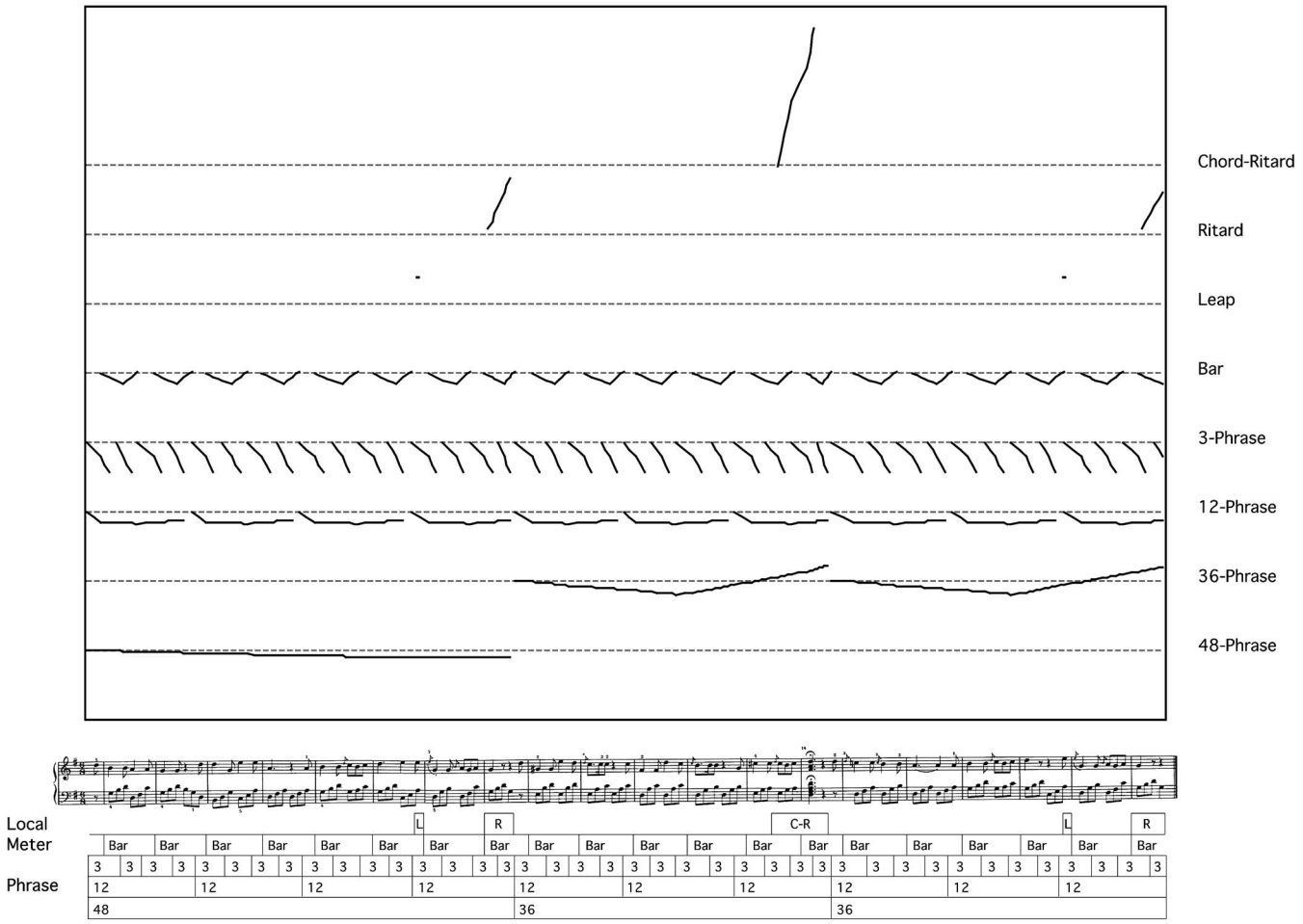


FIG. 5. The decomposition of the expressive profiles: the *x* axis shows score position, the *y* axis the parameter values (in eighth note IOIs in seconds). The names of the structural units are displayed to the right of each profile. Score annotations as in Fig. 4. Note that the *x* axis is warped to align with the musical notation, distorting the canonically regular shape of the profiles.

tempo. The 12-phrase's last breakpoint was also clamped to zero. Thus all profiles reflect timing deviations relative to the tempo specified in the intercept parameter.

D. Example analyses

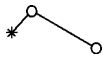
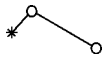
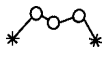
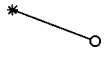
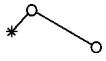
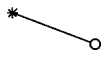
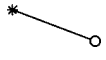
1. Structural decomposition at medium tempo

Given the structural annotation described above, the method fits the prediction to the medium tempo data quite tightly, as is shown in Fig. 4, explaining about 95% of the variance. This is a good fit considering the number of data points (122) and free parameters (13). Fitting the structure to the five individual performances rather than their average delivered only slightly worse fits, ranging between 92% and 95%, indicating that the averaging process indeed removed a small amount of unsystematic noise, possibly motor noise. Moreover, to check how far the model generalizes over repetitions, parameter values obtained in the fit to one performance were applied to all others. This full cross-validation gave again only slightly worse results as on average 92% of the variance was explained: repeated performances are so much alike that a single model generalizes over them. This also means that the successful fit cannot be attributed to overfitting the data: regularities in expressive timing are discovered by the model in a robust manner.

As well as accounting for local expressive timing, such as the pattern associated with the smallest phrase-structure units (three eighth notes long), the model captures more global aspects, such as the ritardandi at the ends of the large-scale phrases. There are some points at which the fit is less good, but a decision was made to accept a trade-off between fit and number of free parameters. It might have been possible to improve the fit at the local phrase level by associating different phrases of the same length with different parameters, but this would have significantly decreased the extent to which the model constrains the data. Similarly, it might have been possible to achieve a better fit with a somewhat more irregular phrase structure, one which allows, for example, for both upbeat and afterbeat phrasings. To test this hypothesis a fit was calculated for a near identical structure to that shown in Fig. 4. The difference here was that the continuous 3-phrase units were replaced with the following pattern: 3342 3342 3333 3333 3342 3342 3333 3333 3333 3333. The fit was near identical (to three decimal places), but requires an extra three free parameters.

However, the decomposition into expressive components that the method achieved is more interesting than the overall fit. Figure 5 shows what (and where) the various components contribute to the total profile.

TABLE I. Structural units and their duration, shape (* is a breakpoint clamped to zero, o is a breakpoint controlled by a free parameter), and number of (free) parameters.

Kind	Name	Description	Extent in 8th notes	Shape	Parameters (free)
phrase	48-Phrase	Opening phrase at same hierarchical level as 36-phrase. Allows for acceleration or deceleration towards and away from central breakpoint.	48		3 (2)
	36-Phrase	Two equal length phrases at same level as 48-phrase. Allows for acceleration or deceleration towards and away from central breakpoint.	36		3 (2)
	12-Phrase	Sub-divides the 48- and 36- phrases. Contains extra breakpoints to allow for agogic accent for anacruses and micropause for last event.	12		5 (3)
	3-Phrase	Lowest level in grouping structure. Allows for acceleration or deceleration within this span.	3		2 (1)
meter	Bar	Profile reflecting the 6/8 metre, with upbeat, and incomplete final bar. Allows for acceleration or deceleration towards and away from breakpoint.	6		3 (2)
local	Leap	Delayed note preceding a grace note to a downwards leap. Only two occurrences.	1		1 (1)
	Chord-Ritard	Slowing towards the fermata.	8		2 (1)
	Ritard	Slowing down at end of first and last long phrase.	5		2 (1)

The results of the method can now be used to detail the links between the different musical structures and the expressive timing and to learn about the interpretation of this specific piece. The most important structural unit was a large slowing down (chord-ritard), explaining three quarters of the variance. The decomposition reveals that in addition to this deceleration towards the fermata, there are less extreme ritardandi (ritard) at the ends of each major phrase, and a repeated acceleration is present over each three eighth-note unit (3-phrase). Over the beginning of the piece (48-phrase) the pianist accelerates gradually, and in each of the subsequent phrases (36-phrase) he follows a schematic acceleration-deceleration profile [familiar from work such as Todd (1992)]. A local lengthening occurs for each note preceding the downward leaps in the melody (leap), marking this distinctive feature. At an intermediate level in the phrase structure (12-phrase) there is an agogic accent on the first event (the upbeat) followed by a slight acceleration deceleration profile. Lastly, the metrical structure is marked in a highly schematic manner, with a pattern of linear acceleration/deceleration across the six beats, rather than a marking of particular beat strengths according to their hierarchical importance (such as described in Palmer and Krumhansl, 1990; Parncutt, 1994).

It has to be stressed that the method is well suited to exploratory data analysis: trying out different structural descriptions and checking out how far they help the fit to the data. In arriving at this successful structural description a number of alternatives were tried. Small increases in the goodness of fit could be achieved by adding parameters associated with additional features, but these increases were regarded as too expensive. For example, the addition of two subphrases of 24 eighth-note units duration within 48-phrase adds two free parameters but only improves the fit by less than 1%. Other structural descriptions that were tested but failed to improve the results were phrases of six and nine units. The most critical improvements in fit/parameter ratio were achieved by creating separate profiles for the ritardandi. This allows for the relatively extreme tempo change at and before the fermata. Table II shows the size of each profile, measured as relative standard deviation, the significance of their contributions, and the amount of variance that each explains. Though some effects and contributions are small, all profiles contribute significantly, and the significance of some contributions is extremely high. Note that these fits arose using 13 parameters, predicting 122 data points. Because some profiles only contribute to a time segment of the data, the amount of variance explained in the whole performance

TABLE II. Explained variance, significance of fit, size (in proportion to the sd of the observed performance), explained variance in stepwise residues, and the number of parameters of each structural component and the complete model.

Profile	r^2	$p<10^{\wedge}$	Proportional sd	Stepwise r^2	Parameters (free)
Chord ritard	0.79	−52	0.82	0.79	2 (1)
3 Phrase	0.08	−21	0.28	0.45	2 (1)
12 Phrase	0.05	−2	0.08	0.18	5 (3)
Ritard	0.01	−11	0.18	0.17	2 (1)
Bar	0.02	−4	0.09	0.15	3 (2)
36 Phrase	0.34	−5	0.14	0.17	3 (2)
Leap	0.02	−3	0.09	0.11	1 (1)
48 Phrase	0.01	−2	0.07	0.07	3 (2)
Full model	0.95	−65	0.98	Nil	21 (13)

TABLE III. The explained variance and the significance of the contribution of the structural units in each tempo.

	Fast		Medium		Slow	
Profile	r^2	$p<10^{\wedge}$	r^2	$p<10^{\wedge}$	r^2	$p<10^{\wedge}$
Chord-ritard	0.49	−17	0.79	−52	0.65	−25
3-Phrase	0.20	−13	0.08	−21	0.07	−9
12-Phrase			0.05	−2		
Ritard	0.01	−2	0.01	−11	0.02	−5
Bar			0.02	−4		
36-Phrase	0.36	−3	0.34	−5	0.43	−4
Leap			0.02	−3		
48-Phrase			0.01	−2		
Full model	0.80	−33	0.95	−65	0.84	−39

is not a very good indication of their relative importance. Otherwise one would be tempted, for example, to be satisfied with the huge contribution of the local chord-ritard, which by itself leaves the expressive timing of most of the piece undefined. In contrast, one would be tempted to underestimate the contribution of the 48-phrase to the beginning section, as the correlations are calculated over the whole piece. However, a somewhat more fair evaluation can be obtained by calculating stepwise residues and reporting the variance explained by each subsequent profile in the corresponding residue. For this, profiles are ordered by explained variance in the remaining residue. This is shown by the fifth column of Table II and demonstrates that some profiles with a small contribution to the overall model are quite good predictors after some other profiles have already been taken care of.

2. Discussion

The decomposition found supports a structural analysis that includes both global features, such as long- and short-term periodicities in metrical and phrase structure, and local features, such as the pianist's response to the fermata. Where periodic structural features are present, a model predicting that the corresponding expression will be fairly similar at each repetition succeeds in predicting this pianist's average behavior very well. Although the large and expected effect of the fermata is highlighted, almost half of the remaining variance can be explained by a repeated pattern of expressive timing at the level of the smallest subphrase (3-phrase). However, all the other profiles account for significant proportions of the variance as well, and the method helps highlight the components that make up the performance.

Two aspects here are worth commenting further on. First, the expressive timing does seem to reflect a concern with local aspects of the musical structure at the expense of more global tempo rubato over longer structural spans. This would be in line with a less "romantic" interpretation of this piece, which is, after all, from the classical period. Second, it is interesting to note that the method allows one to disambiguate between the metrical and short-duration phrase structures, which are out of phase by one eighth-note unit, but multiples of one another. Although the local phrasing accounts for much of the variance in expressive timing, the metrical structure improves the fit still further and manages to predict the expressive timing significantly on its own. Much of the expressive timing follows patterns often observed (see, e.g., Palmer, 1989) or predicted (see Todd, 1992) for classical-romantic repertoire, but our approach allows one to easily observe how such patterns are applied in a nonconsistent manner. There is evidence for both acceleration towards the middle of the phrase here, but also acceleration through such intermediate points towards a phrase boundary. This argues against any model that applies a fixed rule to similar structures across a whole performance. If performers select and combine expressive strategies in a piecemeal manner, inflexible rules cannot capture the decisions that lead to such flexibility. In other words, a model needs to cope with both the extent to which a rule is applied (its weight), but also must account for which rule to apply to any given structure. A choice between accelerating towards a goal or marking it with a gradual deceleration would be a challenging one to simulate, especially where multiple rules may be operating on the same data-points.

3. Application to the analysis of tempo and timing

As shown above, there is evidence that the performer did not maintain proportional timing across the three tempo conditions. Applying the same methods and structural analysis as above but to the data for slow and fast instructed tempi result in good fits as well ($R^2=0.80$ for the fast tempo and $R^2=0.84$ for the slow tempo). The fits and patterns for the profiles are quite similar to that for the medium tempo, although the cumulative fit is not as good (cf. $R^2=0.95$). Table III lists the amount of variance the individual profiles explain in the different tempo conditions.

To check if our choice of tempi was reasonable, the model was run on the data of all nine tempi obtained in Windsor *et al.* (2001), averaged over repetitions. The best fit ($R^2=0.95$) was indeed obtained with the medium tempo, the worst with the fastest ($R^2=0.79$) and the slowest ($R^2=0.84$). The second-fastest and second-slowest tempo, and all tempi in between, allowed the model to explain 92% of the variance or more. This may indicate that at the extreme tempi the possibility to control the performance reliably starts to break down, but that the model and the structural description hold very well for the largest part of the tempo range. The cross

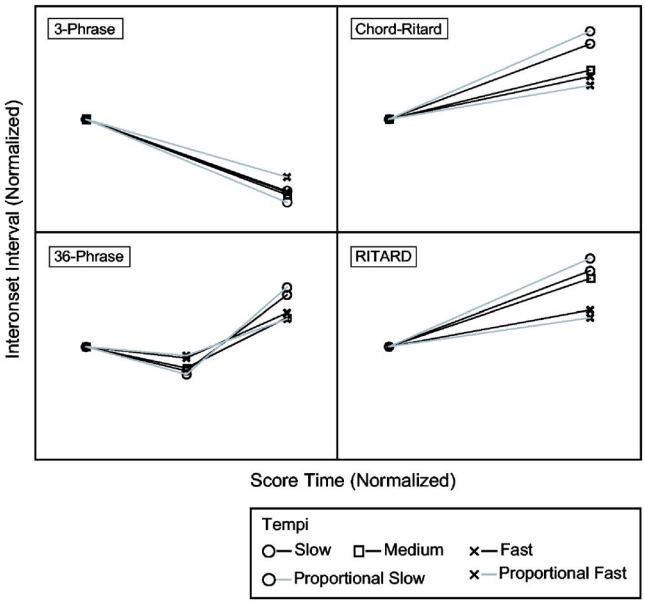


FIG. 6. Panel of the shapes component profiles in the three tempi. Here only one repetition of each profile is shown, and the magnitudes on the x and y axes are normalized for easy comparison of shapes. Reference lines have been added to show proportional tempo predictions for slow and fast tempi taking medium tempo as the baseline.

validation of the model for performances at a specific tempo, with parameters obtained from a performance at another tempo, resulted on average still in 88% explained variance. This again shows that we do not overfit: the model generalizes to a certain extent even across tempi and captures performance regularities in a robust manner. It is, however, interesting to investigate further which profiles adapt in a tempo-specific manner and which do not.

The full models at the fast and the slow tempo become simpler, as some profiles (the weakest in the medium tempo) fall below significance. An advantage of the DISSECT method is that both the relevance and the shape of the profiles can be taken into consideration as they adapt to various (tempo) conditions. The differences in expressive timing for individual profiles at each tempo are shown in Fig. 6, showing only the contributions that are significant at all tempi. In this graph the horizontal and vertical axes are normalized, losing the size of the effects and focusing only on the shape of the profiles, as they adapt to tempo.

The deviations from proportionality for each structural component can now be clearly seen in Fig. 6, as reference lines for proportional invariance are added. The profile for 3-phrase does not scale at all across tempi, while the other profiles scale with tempo, though a bit less than truly proportionally. Such analyses of the scaling behavior of individual profiles with regard to tempo could be practically applicable to the design of a "smart tempo knob" that would adapt performance timing to a set tempo, just like a human pianist would.

4. Discussion

It has been demonstrated that, although highly correlated, the performances at the different instructed tempi are not proportionally invariant. It is therefore useful to be able to show how the differences in expressive timing at the three tempi are related to the structure. Here, the differences can be attributed to subtle changes in the expression of the structural components, partly related to their size, the proportion allocated to the components in relation to others, and partly related to their shape, to the nonproportional scaling of the expression itself (3-phrase). Whether unconsciously or consciously, the pianist has reduced the contribution of the fermata, especially at the fastest tempo, and increased the relative contribution of the shortest phrase unit (3-phrase) by keeping its absolute size invariant. Without the decomposition such aspects of expressive timing are almost impossible to disentangle, and one is left with only qualitative and inductive comparisons of three almost identical tempo profiles. It is possible that the pianist's lessening of attention to the fermata and its preceding onsets and greater accentuation of low-level phrase structure might reflect a less "romantic" interpretation at the faster tempo, which, given the association of faster and less flexible tempo with more 'classical' styles of playing (see Hudson (1994), although see also Bowen (1996) for evidence that such changes are far from systematic), would not be unwarranted.

IV. GENERAL DISCUSSION

The approach to expressive timing described here may seem remarkably underspecified compared to others. Its use of free parameters allows for extremely good fits given a sensible structure, and it could be argued that this is a conceptual weakness. However, this is precisely what allows it to reveal the detailed relationships between structure and expressive timing in these performances. Other theories in this area specify explicit rules and shapes of the expressive profiles, either through experimentation (see, e.g., Sundberg, 1988) or machine-learning (as in Widmer and Tobudic, 2003). The approach developed here reflects a desire to learn these shapes in a somewhat more data-driven way, though of course the structural description is not inferred from the data, it is the given top-down assumption upon which the analysis builds. The examples reported above show the potential of this approach. This methodological concern has a matching theoretical counterpart. It is by no means a logical necessity that rule-following behavior is underpinned by symbolic rules. Even if an aspect of human behavior is, to an extent, systematic, this does not mean that it is rule governed. If expressive timing and musical structure are related, it is not the case that such a relationship need be governed by a mentally encoded set of production rules. Instead, it may be the case that performers learn somewhat similar ways of mapping structure to timing, but that these mappings are both flexible and interchangeable: a performer might choose to accelerate *or* decelerate towards the end of a phrase (see Palmer, 1989 and above) and she or he needs to decide how to combine different kinds of expression both within and between performances of the same and different pieces. These choices may be highly individual or related to stylistic or interpretative differences. Given that this is the case, models proposing a generalized set of rules mapping structure onto expression may only reveal what is least interesting in

musical performance (the way most performers play most of the time) and what is needed is a set of modeling tools that can reveal systematic patterns in performance expression in individual performances, not just those that are shared between many performances. This paper describes a methodology that focuses on this level of explanation, and yet may also discover general properties of expression which might not be captured by stricter models: not only did the parameter values for one performance generalize to an extent to repeated performances, but also across tempo conditions. Not only does this inform us about the regularities in musical performance, it also proves that we are not overfitting the data (explaining nonsystematic features of the training data with high accuracy).

Of course, the model and application presented here require further development. At present the model has only been applied to performances of a single piece by a single performer. A future aim is to show how this approach can illuminate the systematic yet individual nature of expressive timing and dynamics in performances of other pieces. Further research will show if the same approach can deal with nonlinear profile shapes (like the parabola). Another intriguing question, to be addressed in subsequent work, is if the use of a mixed additive-multiplicative model is better than the present linear version. It would separate tempo factors (which combine multiplicatively) and time shifts (which combine additively). Moreover, in principle it should be possible to generate possible structural descriptions automatically (within reasonable constraints regarding meter and phrases) and search for a compact description that explains the data well. This would make the method even more data driven and automatic.

We would argue that editing musical expression by remixing expressive profiles to create a new performance with a different expressive balance and focus would enable the same extensive and parametric control in the area of interpretation that is already commonplace for sound synthesis, filtering, and spatialization. For such practical applications an expression synthesizer has been developed as a companion to the DISSECT analysis method. The synthesizer mixes a new performance with edited expression using the profiles yielded by DISSECT and a set of weights. These weights control the extent to which a specific expressive component is present in the output. Informally, the results sound quite promising: for example, performances with exaggerated bar timing or muted final ritard timing sound to us as if they have been played by a human performer who was instructed to play in that way. A more elaborate discussion and a demonstration of the expression mixer is available at the MMM website (⟨www.nici.ru.nl/mmm⟩ under demos). Generating expression profiles with muted, exaggerated, or otherwise perturbed components provides a rich domain of stimuli that can be used to probe perceptual and motor processes (see, e.g., Clarke, 1993; Clarke and Windsor, 2000); greater and more detailed control of the parameters in such experiments would allow researchers to modify only certain aspects of expression while leaving others invariant.

V. CONCLUSIONS

This paper has presented an approach to the decomposition of expressive timing which can be thought of as a generalized model of the mappings between structure and expression that have been empirically observed since the time of Seashore (1938). We have shown how this approach can independently predict different structural components in expression, how it is sensitive to subtle changes in interpretation by a single performer, and how it provides standard and interpretable estimates of fit. Although the decomposition method only makes a few assumptions about the "rules" which map structure onto expression, it provides a sensitive and systematic method for gaining insight into what constraints operate in the domain of musical expression, a topic which, despite concerted effort and much excellent research, still seems to pose many questions. Researchers know something about what performances have in common and how they differ in gross terms (see, e.g., Repp, 1992) and are sometimes able to model some of these generalities (see, e.g., Todd, 1992). The focus here is on the subtleties of expression in a single performance and how these change under different performance conditions, and, as we have shown, such subtleties can be effectively highlighted if one systematically decomposes the expressive signal into multiple components.

ACKNOWLEDGMENTS

The authors would like to acknowledge the advice and help that they received during the course of this research from Eric Maris, Henkjan Honing, Renee Timmers, Makiko Sadakata, Diana Deutsch, and from two anonymous reviewers. This work was funded by the Netherlands Organisation for Scientific Research (NWO); the Faculty of Social Sciences, Radboud University; the Faculty of Performance, Visual Arts and Communications, University of Leeds; and a fellowship in cognitive sciences from the French Ministry of Education and Research and a PECA (Perception Et Cognition Auditive) travel fellowship.

Bowen, J. A. (**1996**). "*Tempo* Duration & Flexibility: Techniques in the Analysis of Performance," J. Music. Res. **16**(2), 111–156.

Chaffin, R., and Imreh, G. (**2002**). "Practicing perfection: Piano performance as expert memory," Psychol. Sci. **13**(4), 342–349.

Clarke, E. F. (**1988**). "Generative principles in music performance," in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, edited by J. A. Sloboda (Clarendon, Oxford), pp. 1–26.

Clarke, E. F. (**1993**). "Imitating and evaluating real and transformed musical performances," Music Percept. **10**(3), 317–341.

Clarke, E. F., and Windsor, W. L. (**2000**). "Real and Simulated Expression: a Listening Study," Music Percept. **17**(3), 1–37.

Clynes, M. (**1983**). "Expressive microstructure in music, linked to living qualities," in *Studies of Music Performance*, edited by J. Sundberg (Royal Swedish Academy of Music, Stockholm), pp. 76–181.

Desain, P., and Honing, H. (**1994**). "Does expressive timing in music performance scale proportionally with tempo?" Psychol. Res. **56**, 285–292.

Desain, P., and Honing, H. (**1997**). "Structural Expression Component Theory (SECT), and a method for decomposing expression in music performance," in *Proceedings of the Society for Music Perception and Cognition Conference* (MIT, Cambridge), p. 38.

Desain, P., Honing, H., Van Thienen, H., and Windsor, L. (**1998**). "Computational Modeling of Music Cognition: Problem or Solution?" Music Percept. **16**(1), 151–166.

Dixon, S., Goebl, W., and Widmer, G. (**2002**). "Real time tracking and visualisation of musical expression," in *Proceedings of the Second International Conference on Music and Artificial Intelligence*, edited by C. Anagnostopoulou, M. Ferrand, and A. Smaill (Springer, Berlin), pp. 58–68.

Friberg, A. (**1991**). "Generative Rules for Music Performance: A Formal Description of a Rule System," Comput. Music J. **15**(2), 56–71.

Friberg, A. (**1995**). "Matching the rule parameters of Phrase arch to performances of 'Träumerei': a preliminary study," in *Proceedings of KTH Symposium on Grammars for Music Performance* (KTH, Stockholm).

Gentner, D. R. (**1987**). "Timing of Skilled Motor Performance: Tests of the Proportional Duration Model," Psychol. Rev. **94**(2), 255–276.

Honing, H. (**1990**). "POCO: an environment for analysing, modifying, and generating expression in music," in *Proceedings of the 1990 International Computer Music Conference* (International Computer Music, San Francisco), pp. 364–368.

Honing, H. (**1992**). "Expresso, a strong and small editor for expression," in *Proceedings of the 1992 International Computer Music Conference* (International Computer Music, San Francisco), pp. 215–218.

Hudson, R. (**1994**). *Stolen Time: The History of Tempo Rubato* (Clarendon, Oxford).

Palmer, C. (**1989**). "Mapping musical thought to musical performance," J. Exp. Psychol. Hum. Percept. Perform. **15**(12), 331–346.

Palmer, C. (**1997**). "Music Performance," Annu. Rev. Psychol. **48**, 115–138.

Palmer, C., and Krumhansl, C. L. (**1990**). "Mental representations for musical meter," J. Exp. Psychol. Hum. Percept. Perform. **16**(4), 728–741.

Parncutt, R. (**1994**). "A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms," Music Percept. **11**(4), 409–464.

Penel, A. (**2000**). "Variations temporelles dans l'interprétation musicale: processus perceptifs et cognitifs," unpublished doctoral dissertation, Université Paris 6. Available from the third author on request.

Penel, A., and Drake, C. (**1998**). "Sources of timing variations in music performance: A psychological segmentation model," Psychol. Res. **61**, 12–32.

Penel, A., Desain, P., Maris, E., and Windsor, W. L. (**1999**). "A decomposition model of expressive timing," in *Proceedings of the 1999 SMPC* Evanston, IL, p. 21.

Repp, B. H. (**1992**). "Diversity and commonality in music performance—an analysis of timing microstructure in Schumann's Traumerei," J. Acoust. Soc. Am. **92**, 2546–2568.

Repp, B. H. (**1994**). "Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study," Psychol. Res. **56**, 269–284.

Schmidt, R. A. (**1985**). "The search for invariance in skilled movement behavior," Res. Q. Exerc Sport **56**(2), 188–200.

Seashore, C. E. (**1938**). *Psychology of Music* (McGraw-Hill, New York).

Shaffer, L. H. (**1981**). "Performances of Chopin, Bach and Bartok: studies in motor programming," Cognit Psychol. **13**, 326–376.

Sundberg, J. (**1988**). "Computer synthesis of music performance," in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, edited by J. A. Sloboda (Clarendon, Oxford).

Sundberg, J., Friberg, A., and Bresin, R. (**2003**). "Attempts to Reproduce a Pianist's Expressive Timing with Director Musices Performance Rules," J. New Music Res. **32**(3), 317–325.

Thompson, W. F., and Cuddy, L. L. (**1997**). "Music performance and the perception of key," J. Exp. Psychol. Hum. Percept. Perform. **23**(1), 116–135.

Todd, N. P. (**1985**). "A model of expressive timing in tonal music," Music Percept. **3**, 33–58.

Todd, N. P. (**1992**). "The dynamics of dynamics: A model of musical expression," J. Acoust. Soc. Am. **91**, 3540–3550.

Todd, N. P. McA. (**1995**). "The kinematics of musical expression," J. Acoust. Soc. Am. **97**, 1940–1949.

Widmer, G., and Tobudic, A. (**2003**). "Playing Mozart by Analogy: Learning Multi-level Timing and Dynamics Strategies," J. New Music Res. **32**(3), 259–268.

Windsor, W. L., and Clarke, E. F. (**1997**). "Expressive timing and dynamics in real and artificial musical performances: using an algorithm as an analytical tool," Music Percept. **15**(2), 127–152.

Windsor, W. L., Aarts, R., Desain, P., Heijink, H., and Timmers, R. (**2001**). "The timing of grace notes in skilled musical performance at different tempi: a preliminary case study," Psychol. Music **29**, 149–169.

Zanon, P., and De Poli, G. (**2003a**). "Time-varying estimation of parameters in rule systems for music performance," J. New Music Res. **32**(3), 295–316.

Zanon, P., and De Poli, G. (**2003b**). "Estimation of parameters in rule systems for expressive rendering in musical performance," Comput. Music J. **27**(1), 29–46.

Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I. Experiment[a)]

Alex Z. Tarnopolsky
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

Neville H. Fletcher
School of Physics, University of New South Wales, Sydney NSW 2052, Australia and Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

Lloyd C. L. Hollenberg
School of Physics, University of Melbourne, Melbourne, Vic 3010, Australia

Benjamin D. Lange, John Smith, and Joe Wolfe[b)]
School of Physics, University of New South Wales, Sydney NSW 2052, Australia

(Received 8 August 2005; accepted 8 November 2005)

The didjeridu, or yidaki, is a simple tube about 1.5 m long, played with the lips, as in a tuba, but mostly producing just a tonal, rhythmic drone sound. The acoustic impedance spectra of performers' vocal tracts were measured while they played and compared with the radiated sound spectra. When the tongue is close to the hard palate, the vocal tract impedance has several maxima in the range 1–3 kHz. These maxima, if sufficiently large, produce minima in the spectral envelope of the sound because the corresponding frequency components of acoustic current in the flow entering the instrument are small. In the ranges between the impedance maxima, the lower impedance of the tract allows relatively large acoustic current components that correspond to strong formants in the radiated sound. Broad, weak formants can also be observed when groups of even or odd harmonics coincide with bore resonances. Schlieren photographs of the jet entering the instrument and high speed video images of the player's lips show that the lips are closed for about half of each cycle, thus generating high levels of upper harmonics of the lip frequency. Examples of the spectra of "circular breathing" and combined playing and vocalization are shown. [DOI: 10.1121/1.2146089]

PACS number(s): 43.75.Fg, 43.75.Yy, 43.72.Ct [DD] Pages: 1194–1204

I. INTRODUCTION

The word "didjeridu" (or "didgeridoo" in the popular literature) is an onomatopoeic Western name for a traditional instrument played in parts of Northern Australia and known to the Yolngu people of Arnhem Land as the yidaki. The Yolngu name "yidaki" will be used throughout this paper. It is unusual among wind instruments in that the pitch is only rarely varied: the interest in performance lies in spectacular, rhythmic variations in timbre, which are produced by the player's vocal tract. It is played using "circular breathing" to produce an uninterrupted sound: the player, traditionally a man, fills his cheeks and uses this reservoir to continue to play while simultaneously inhaling quickly through the nose (and bypassing the mouth at the soft palate) to refill the lungs with air. The differences between timbres produced by playing using the mouth cavity alone, while inhaling, and those produced using the complete tract, while exhaling, are usually unavoidable and are incorporated into the rhythmic variation in timbre that is idiomatic for the instrument and that gives it its Western name. Different tongue positions have a strong effect on the sound spectrum. Sound files illustrating these effects are given at www.phys.unsw.edu.au/~jw/yidakididjeridu.html.

The yidaki is a member of the lip valve family. In this family, the playing frequency is usually close to that of one of the maxima in the impedance spectrum of the bore. The effect of variations in the player's vocal tract upon orchestral lip valve instruments is usual modest, because of their narrow bore and the shape of the mouthpiece, whereas in the yidaki it is the preeminent musical feature. The yidaki is therefore an ideal instrument in which to study the interaction among vocal tract, lips, and instrument.

Previous studies of the acoustics of the yidaki have considered the lip motion (Wiggins, 1988), the lip-bore interaction (Fletcher, 1983, 1996), numerical modeling of the lip motion (Hollenberg, 2000), the linear acoustics of the instrument (Amir and Alon, 2001; Amir, 2004), and the acoustics of the vocal tract of players miming playing (Fletcher *et al.*, 2001). A very brief report covering work related to that reported here has been given previously (Tarnopolsky *et al.*, 2005). That report used an instrument made in the traditional manner, and therefore of unknown geometry.

[a)]A Brief Communication reporting related studies has been published in Nature (Tamopolsky et al, 2005).

[b)]Author to whom correspondence should be addressed. Electronic mail: j.wolfe@unsw.edu.au

In what is now a standard model, Backus (1985) proposed that the acoustic impedance of the bore of a wind instrument Z_{inst} and that of the tract Z_{tract} act in series on the valve and on the air flow through it. An understanding of the interaction between the instrument and the vocal tract thus requires detailed knowledge of the acoustic impedance of the instrument, the acoustic impedance of the vocal tract, the vibratory behavior of the lip valve, the air jet entering the instrument, and their interaction.

It is relatively easy to determine the impedance of the instrument, particularly if only the maxima are important. However, the impedance of the vocal tract is much harder to measure, particularly during playing. The vibrating lips generate a sound signal that is transmitted both into the instrument and into the mouth. Consequently, the sound level inside the mouth of the player is very high. This makes it difficult to make measurements of the acoustic impedance of the vocal tract of someone playing the yidaki. In the past, we have made measurements of that impedance while players mimed playing (Fletcher *et al.*, 2001). However, producing a given mouth configuration in the absence of audible feedback is difficult, and it is not clear that yidaki players are capable of miming reliably, perhaps particularly with regard to the aperture of the glottis, of which most people are not conscious.

In this study, we report the development of a system that allows the measurement of the acoustic impedance of the vocal tract, just inside the player's lips, while he is playing the yidaki. We compare this with the spectrum of the sound produced. We also report the motion of the player's lips, using high-speed photography. These sets of observations are used to test a simple model that explains how the acoustic impedance of the tract affects the spectral envelope of the sound produced. We also analyze other features of idiomatic playing: "circular breathing" and vocalization.

A. The yidaki (didjeridu)

Traditionally, the material for a yidaki is selected by tapping suitably sized tree trunks to find one whose interior has been eaten by termites to provide a suitable central bore. It is cut to a desired length, the bore is cleaned and sometimes shaped further, and a ring of beeswax is fitted to the smaller end to make a comfortable seal for the player's lips. The outside is sometimes painted with traditional designs of cultural significance. The instrument is typically 1.2 to 1.5 m long (different cultural groups have different styles) and has an irregular bore, which is usually somewhat flared from about 30 to 50 mm at the blowing end to about 40 to 150 mm at the open end. Sealed at the lip end and open at the other, its lowest resonance is typically 50 to 80 Hz. Orchestral wind instruments usually have several bore resonances whose frequencies fall in harmonic ratios. Because of its shape, this is not usually the case for the yidaki: it is usually neither a cylinder nor a nearly complete cone and its resonances form a "stretched" quasi-harmonic series (Fletcher, 1996). Consequently, harmonics of the note being played only sometimes coincide with a bore resonance of the instrument. Higher or "overblown" notes near the frequencies of the second or third bore resonances may be sounded briefly for contrast. This is not usual in traditional playing in Western Arnhem Land, though it is used in the East.

Because the bore is typically 30 to 50 mm in diameter at the smaller end, its characteristic impedance is lower by an order of magnitude than that of most of the members of the lip-driven musical instrument family such as the horn or trumpet. This and the rather rough walls of the bore imply that the magnitudes of the maxima in the yidaki's impedance spectrum are rather lower than those of other members of the wind instrument family. Some consequences have been discussed in previous papers (Fletcher, 1996; Amir and Alon, 2001; Fletcher *et al.*, 2001; Caussé *et al.*, 2004).

Because the shape of the bore is largely determined by termites and the shape of the tree trunk, the variation among these instruments is great. The purpose of the current study is to investigate the principles of operation, rather than the effects of different instrumental geometries (which is the subject of another study). For that reason, and to facilitate reproduction of the results reported here, two model instruments were used. For acoustical measurements, we used a cylindrical PVC pipe. For optical measurements, we used a pipe made of plexiglass with a square cross section. Experienced players reported that both model instruments played moderately well. Indeed instruments with a constant cross section, usually made of PVC pipe, are occasionally used in nontraditional musical contexts, particularly when a given pitch is required in order to play with other instruments.

B. Vocal tract-instrument interaction

There are a number of reports on the effects of the vocal tract on the sound on orchestral wind instruments (Elliot and Bowsher, 1982; Clinch *et al.*, 1982; Wolfe *et al.*, 2003), but the effects in such instruments are modest in comparison with those in the yidaki. These orchestral instruments have a narrow constriction in the mouthpiece and a smooth bore, which is typically only several mm in the mouthpiece. These features give the instruments an impedance spectrum with a series of maxima whose values exceed considerably those of the vocal tract. Consequently, there is only modest coupling between the two resonators (the vocal tract and the bore of the instrument). While the effect of the tract on timbre of orchestral instruments is large enough to interest composers (e.g., Berio, 1966; Erikson, 1969), it is small compared to the striking effects of the vocal tract on the timbre of the yidaki. In a previous paper (Fletcher *et al.*, 2001) we reported sound spectra, vocal tract configurations, and the impedance spectra of players miming the playing of the yidaki. However, these were not measured simultaneously during playing, so we were then unable to make quantitative comparisons among them.

C. Measurement of the vocal tract impedance during performance

The magnitude of the vocal tract effect in the yidaki makes it an ideal instrument upon which to study vocal tract

effects in general. For this purpose, we have adapted an impedance spectrometer described previously (Epps *et al.*, 1997) to allow us to make impedance measurements using an impedance probe placed just inside the player's lips, while he is playing. This situation requires several practical compromises. The sound due to the playing has comparable levels in the mouth and in the instrument. As this is "noise" for the purposes of measurement of the impedance spectrum, the signal-to-noise ratio is low. There is the further complication of a humid environment, which means that water-resistant or disposable microphones must be used. On the other hand, this study is concerned with relating the spectral envelope of the sound produced to the overall features of the impedance spectrum. Consequently, high-precision calibrated microphones are not required.

II. MATERIALS AND METHODS

A. Yidakis

For acoustic measurements, a "model yidaki" was made of cylindrical PVC pipe, length 1210 mm and inner diameter 30 mm. (It is referred to as "the pipe" or "the instrument" below.) For the optical measurements of the lip motion and for the flow visualization, a pipe with square cross section was made of plexiglass with glass panels for the optical pathway. It is 1220 mm long and the internal width is 38 mm.

B. Players

One of the players, BL, traditional name Wilamara, is a member of the Mara people of Roper River in Northern Australia, where he learned to play yidaki in the traditional style. LH is an Australian of European cultural background who has been playing the yidaki for 8 years. AT is an Australian of European cultural background who learned to play yidaki for the purposes of this study. LH's usual playing style has the instrument displaced laterally from the center of the lips. Neither he nor AT had trouble adapting to the presence of the impedance probe behind the lips. BL, who has played for the longest time and whose lip-instrument position is symmetrical, found the impedance probe disruptive, particularly for the high tongue position.

Players were asked to produce three different mouth configurations for recordings. One is called a high tongue drone (hereafter "high tongue"): the player holds the tongue close to the hard palate so that there is a constriction in the air passage between the throat and the lips. This produces a strong formant between about 1.5 and 2 kHz, whose frequency and amplitude may be varied by the performer. This sound is very common in yidaki performance. In another configuration, hereafter called "low tongue," the players were asked to play with the tongue low in the mouth and thus no lingual constriction. This configuration produces a sound without a strong formant and is used as a contrast to the high-tongue drone. In the third configuration, players inflated their cheeks and then expelled the air, while inhaling, as described above under "circular breathing." In a different series, they were asked to vocalize at harmonic intervals above the note they were playing on the yidaki.

To measure the static mouth pressure during playing, they were also asked to play with a range of loudness levels while a small tube connected the mouth cavity to a water manometer.

C. Measurements of impedance spectra of the instrument

An impedance spectrometer described previously (Smith *et al.*, 1997; Epps *et al.*, 1997) was adapted for this study. Briefly, a waveform is synthesized from harmonic components, amplified, and input via a loudspeaker and impedance matching horn to a narrow high-impedance tube leading to the item under test. This approximates an ideal source of acoustic current. It is calibrated by connection to a reference impedance, which is an acoustically quasi-infinite cylindrical pipe whose impedance is assumed to be real, frequency independent, and equal to its calculated characteristic impedance. From the coefficients of the spectrum of the measured sound in this calibration stage, a new signal is synthesized to produce a measured spectrum with frequency components of equal amplitude. This is used as the acoustic current source for subsequent measurements and the unknown impedance spectrum is calculated from the pressure components measured in measurement and calibration stages, taking into account the small, parallel admittance of the source. The bore diameter of the yidaki is larger than that of the instruments we have studied previously (Wolfe *et al.*, 2001) and consequently a lower impedance reference was required. The acoustically infinite cylindrical pipe used for calibration in this study had an internal diameter of 26.2 mm and a length of 194 m. Because the first curve in the pipe occurs at 40 m from the spectrometer and because any curves have a radius of 5 m or greater, the effects of reflections from these curves are expected to be negligible and this reference impedance should be purely resistive.

D. Measurements of impedance spectra in the vocal tract

The microphone recording the pressure inside the tract is exposed to high humidity and high steady pressure. For this reason, we used inexpensive electret microphones (Optimus 33-3013), which were replaced when necessary. It was necessary to attenuate the acoustic signal to avoid clipping or harmonic distortion in these microphones. To do this, we used the acoustic divider circuit shown in Fig. 1. At low frequencies, where attenuation is most important, the impedances of both pipes in the divider (including the radiation impedance associated with the open end of the tube) are essentially inertive, and the phase change along the length is small. At higher frequencies, there is a phase shift and a frequency-dependent gain. The microphone therefore records only a fraction of the pressure inside the mouth. This attenuation is inside the calibration loop for the impedance probe, so its frequency and phase response do not affect measurements. The source capillary has an inner diameter of 3.7 mm and the microphone tube an inner diameter of 1.5 mm. Both have a length of 35 mm. Measurements were made at a frequency spacing of 5.383 Hz from 0.2 to 3.0 kHz.

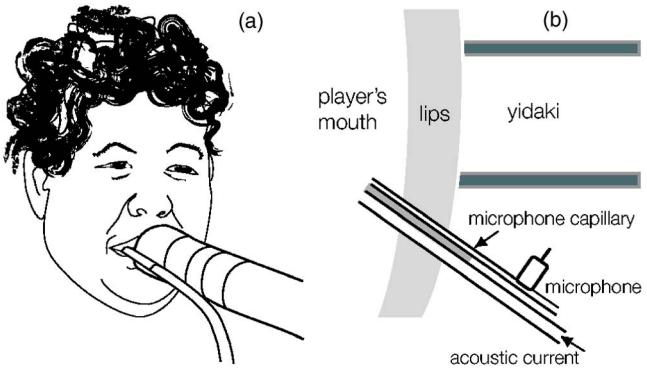


FIG. 1. The technique used to measure the impedance of the vocal tract during performance. The sketch shows the impedance probe inserted into a corner of the mouth. The schematic (top view—only approximately to scale) shows the geometry of the impedance probe and its location in the player's mouth. The microphone capillary is hidden in the figure at the left.

While the acoustic pressure acts upon a significant area of the lips and this determines their vibratory motion, the pressure of concern in the production of formants in the sound is that acting over the opening area of the lips. Measurements of acoustic impedance need to take this geometrical mismatch into consideration, since essentially there is an inertive correction, which may be either positive or negative, involved between this impedance and the plane-wave impedance normally measured in a pipe (Brass and Locke, 1997; Fletcher *et al.*, 2005). The probe used for measurement of the vocal tract impedance, shown in Fig. 1, has a narrow outlet for the acoustic current. The impedance spectrometer is calibrated on a quasi-infinite tube of diameter 26.2 mm, which is comparable with the size of the vocal tract. Consequently, errors due to this effect are small.

These measurements required the impedance probe shown in Fig. 1 (8 mm wide and 5 mm high) to be placed in the mouth during playing. Impedance measurements start after the player gives a signal that he is happy with the tongue position and the sound produced. The player then continues to play on one breath, typically for about 10 s. During this time an impedance measurement is made. A sample of the sound immediately following the impedance measurement is used to obtain samples of the radiated sound uncontaminated by that of the injected measurement signal.

E. Measurements of sound spectra

The output sound was recorded on digital audio tape at 44.1 kHz using an omnidirectional electret microphone, placed on the axis of the yidaki, at a distance of 125 mm from its end.

F. Measurements of the lip motion

The transparent yidaki with square cross section (mentioned above) had two mouthpiece configurations. For measurements of the lip motion, a round hole was cut in one side for the player's lips. To improve the image quality, sections of plexiglass on the side and end were replaced with two glass panels, 100 mm long and equal in width to the yidaki, inserted in the optical path at the mouth end, as shown in Fig. 2. A mirror, mounted vertically on the end at 45°, allowed the single camera to record the plane and lateral image of the player's lips simultaneously. Before each recording, the glass panels were heated with warm dry air to prevent water condensation. A video camera running at 1000 frames per second was used to record the images.

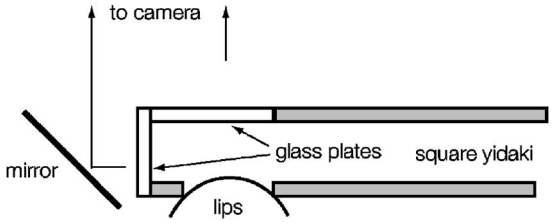


FIG. 2. The geometry of the square yidaki, seen from above, as configured for high speed photography.

G. Flow visualization of the jet motion

Images of the air jet inside the yidaki during playing were achieved using schlieren imaging, a nondestructive optical flow visualization technique, which is described elsewhere (Tarnopolsky and Fletcher, 2004).

This experiment also used the plexiglass yidaki. The schlieren technique depends upon refraction of light rays as they pass through regions with varying refractive index, usually provided by inhomogeneities in density. A curtain of higher density gas was produced by releasing carbon dioxide from a manifold on the outer side into the yidaki through a line of 11 holes of 1.5-mm diameter linking the manifold to the yidaki (see Fig. 3). During playing, the air jet passes through and is contaminated by the curtain of carbon dioxide and thus generates the necessary density gradient.

A light source that produced a single pulse of duration of about 0.2 ms was triggered electronically at a selected phase of the lips' opening (Tarnopolsky *et al.*, 2000). Stroboscopy of a steady, sustained playing gesture was used to obtain images covering one period of the lip's oscillation in time steps of 1 ms. The period of the lip oscillation was 14.3 ms for these experiments.

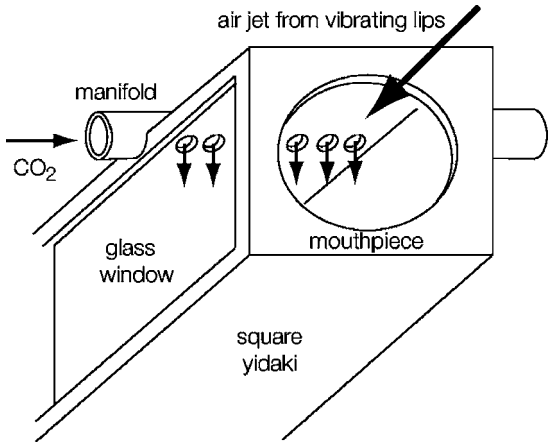


FIG. 3. The geometry of the square yidaki, as configured for flow visualization.

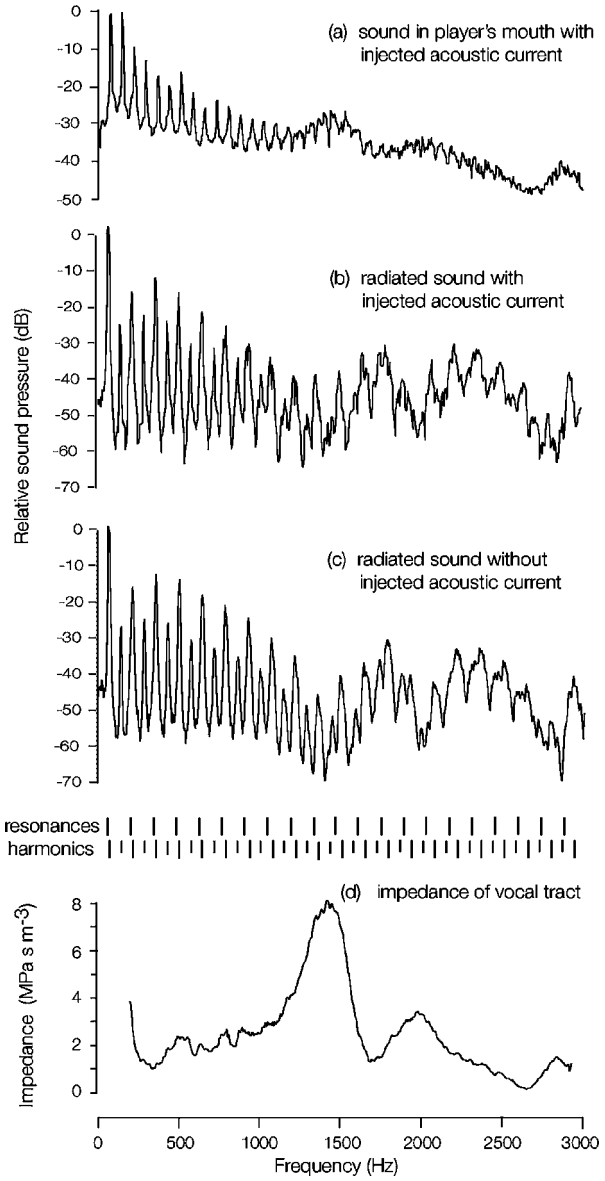


FIG. 4. A figure to illustrate the vocal tract impedance measurement while the subject is playing in the high tongue configuration. (a) shows the spectrum of the sound pressure level measured in the mouth due to both lip vibration and the injected acoustic current. (b) shows the radiated sound spectrum measured simultaneously with (a). (c) shows the radiated sound measured just after the impedance measurement i.e. without the injected acoustic current. (d) shows the impedance of the vocal tract that was derived from (a). The sound pressure levels shown in (a), (b), and (c) are normalized relative to their largest frequency component. The frequencies of the harmonics of the sound and of the resonances of the pipe are also shown with vertical dashes. The odd numbered harmonics are represented by longer dashes.

III. RESULTS AND DISCUSSION

A. Impedance of the player's vocal tract

Figure 4 shows how measurements of the acoustic impedance of the player's tract were made during performance and then processed. Figure 4(a) shows the spectrum of the signal recorded inside the mouth during a typical example of the high tongue configuration. The periodic vibration of the lips at about 70 Hz generates an acoustic signal that interacts with the impedance of the tract to produce a series of harmonics, which are seen at frequencies below about 1.5 kHz. Inside the mouth, the higher resonances of the instrument (some of which lie close to odd harmonics of the lip motion) have only modest influence on the sound in the mouth. Hence, in Fig. 4(a), there is no systematic difference between odd and even harmonics. At frequencies above 1 kHz (the range of interest), the spectrum is increasingly dominated by the response of the vocal tract to the injected acoustic current. Here we see broad peaks or formants at about 1.5, 2.1, and 2.8 kHz. Because the injected current has been calibrated to have flow components with magnitude independent of frequency, these peaks correspond to maxima in the acoustic impedance spectrum of the vocal tract in this configuration.

Figure 4(b) shows the spectrum of the externally radiated sound produced by the yidaki, measured 125 mm from the end of the instrument. This shows the strong harmonics of the instrument's sound. Odd harmonics dominate, because of the impedance matching effects of the transfer function of the closed, cylindrical pipe. The figure shows the frequencies of the harmonics of the sound and those of the impedance maxima in the cylindrical pipe (titled "resonances"), which are approximately at nf_1 where n is an odd integer and f_1 is the frequency of the lowest resonance. (Because of frequency-dependent end effects, these frequencies are not exactly harmonic.) The effects of resonances on harmonics are discussed in more detail later. This spectrum is included to allow comparison of the harmonics measured simultaneously inside and outside the mouth. Above about 1.5 kHz, the spectrum has an increased broadband component. This is the (filtered) sound of the injected acoustic current. Some of the sound injected into the mouth is radiated through the opening lips and the yidaki. Some also leaks through the cladding of the current source directly into the radiation field. Because of this unavoidable contamination of the radiated yidaki sound by the acoustic current used to measure the impedance, all sound spectra shown in subsequent figures were measured immediately following the impedance measurement, during the same, sustained playing gesture. This is also the case in Fig. 4(c).

Figure 4(d) shows the acoustic impedance of the vocal tract during playing. The acoustic impedance was derived from the signal recorded inside the mouth during the injected sound as described above. Frequencies below 200 Hz (the lowest frequency in the injected current) are omitted. To remove the very large signal produced by the vibrating lips, five points centered on each harmonic of the lip frequency up to the 16th have been removed and replaced with a linear interpolation. The resulting data have been smoothed by a linear average over a window of 53.8 Hz and are presented on a linear rather than a logarithmic scale. This process is used hereafter to show Z_{tract} when measured during playing.

Comparing Fig. 4(b) or 4(c) with Fig. 4(d) shows that the peaks in the vocal tract impedance occur at frequencies at which the spectral envelope of the radiated sound has minima. This is considered further, below.

When considering the spectral envelope of Fig. 4(b) or 4(c), one should remember that human hearing sensitivity declines rapidly below about 300 Hz. Consequently, despite their relatively large amplitude, the fundamental and lower

harmonics are not very loud. Further, they vary little during playing. The formants, on the other hand, occur at frequencies in the range of maximum sensitivity of the ear, and they change in response to changes in mouth configuration. It is these formants and the variation in them that contribute most of the interest in yidaki performance.

B. Relationship between the output sound and the impedance of the player's vocal tract

The playing frequency is close to but slightly above that of the first resonance of the pipe, in accordance with Fletcher's (1993) analysis of an "outward swinging door" valve that opens under excess pressure on the upstream side and closes under excess pressure on the downstream side [notated (+, −)]. The signal radiated by this cylindrical yidaki has stronger odd harmonics, especially for low frequencies, where these harmonics fall close to the resonances of the instrument, which are indicated by vertical lines in Fig. 4(c). The broadband component of the radiated signal (largely due to the spectrometer signal leaking through the player's lips into the yidaki) is visible, especially at high frequencies. This broadband spectrum has an envelope that resembles the inverse of the vocal tract impedance, which is discussed below.

The signal in the mouth in Fig. 4(a), which is due to the interaction of the flow through the vibrating lips with the vocal tract, shows no strong difference between even and odd harmonics, because the resonances in the tract are much broader than the frequency differences between the harmonics of the lip vibration.

The vocal tract impedance shows broad peaks at approximately 1.5, 2.1, and 2.8 kHz. A weak peak below 500 Hz is often seen when players mime playing [data not shown, but see Fletcher *et al.* (2001)], but here it is not seen: it is possibly obscured by the strong signal from the vibrating lips.

Comparing the vocal tract impedance spectrum Z_{tract} in Fig. 4(d) with the radiated sound spectrum in Fig. 4(c) one notes that, when Z_{tract} is sufficiently large, the envelope of the radiated sound spectrum is low. This correlation was evident in many such spectra, both from cylindrical pipes and flared yidakis (data not shown). Figure 5 shows the results of independent measurements of the high tongue configuration for the three players described above. The minima and maxima in the spectral envelope of the radiated sound that fell in the range 1.0 to 2.2 kHz were recorded, as were the extrema in the vocal tract impedance measured immediately previously. In Fig. 5, the frequency of each minimum in the spectral envelope of the radiated sound is plotted against that of the nearest maximum in the impedance spectrum (filled symbols) and the frequency of each maximum in the sound envelope is plotted against that of the nearest minimum in the impedance spectrum. The correlation is excellent (the slope is 0.93 and the correlation coefficient is 0.98). The impedance maxima correspond almost exactly to minima in the spectral envelope of the sound, while maxima in the sound spectrum occur on average at frequencies slightly above those of the minimum in the impedance.

Why does a peak in Z_{tract} reduce the level of the radiated sound? The connection is a little obscure and requires careful

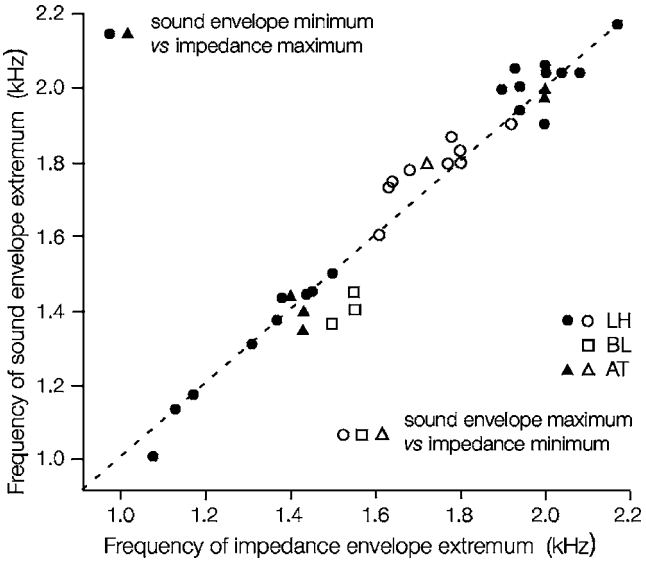


FIG. 5. On this graph, each filled symbol plots the frequency of a minimum in the spectral envelope of the radiated sound against the frequency of the nearest maximum in the impedance spectrum of the vocal tract. Each open symbol plots the frequency of a maximum in the spectral envelope of the radiated sound against the frequency of the nearest minimum in the measured impedance spectrum. No clear impedance maxima were evident for player BL in the measured range. The dashed line is the line of equality.

analysis. For this reason it will be discussed only briefly here, but is treated in detail in our companion theory paper (Fletcher *et al.*, 2006). In most lip-valve instruments, the maxima in Z_{instr} are very much larger than those in Z_{tract}. For the yidaki, in the frequency range of interest (1 to 3 kHz), this is not the case, for two or three reasons. First, the yidaki has a larger cross section than does the mouth with the tongue raised and so it has a relatively small characteristic impedance. Second, in a traditional yidaki, wall losses due to roughness in the bore of a genuine instrument may also be important, though not for the PVC pipe used here. In Fig. 4(d), the peak in Z_{tract} at 1.5 kHz has a value of about 8 MPa s m^{-3} for this configuration when the tongue is raised. The impedance of the pipe has peaks in this frequency range of 3 to 10 MPa s m^{-3}. Consequently, these broad peaks in Z_{tract} give rise to a minimum in the acoustic flow U in the yidaki, at the lips. For any given value of the transfer function between the two ends of the yidaki, a small U at the input yields a small acoustic pressure at the output. Of course, the transfer function of a pipe is a strong function of frequency and has maxima at approximately $f_{\text{pipe}}=(2n+1)c/4L$, where c is the speed of sound, L is the length of the pipe, and n is an integer. (At these frequencies, the pipe is a good impedance transformer to match the relatively low radiation impedance.) But the resonances of the pipe are closely spaced in frequency compared to those of the tract, so that several harmonics of the played sound will fall within a formant produced by a resonance of the vocal tract. This also explains the shape of the broadband component of the radiated spectra, discussed above.

Figure 6 presents another example of a measurement with the high tongue configuration. We propose that the impedance maxima in Z_{tract} measured just inside the lips shown in Figs. 4(d) and 6(c) are due to resonances of the upper

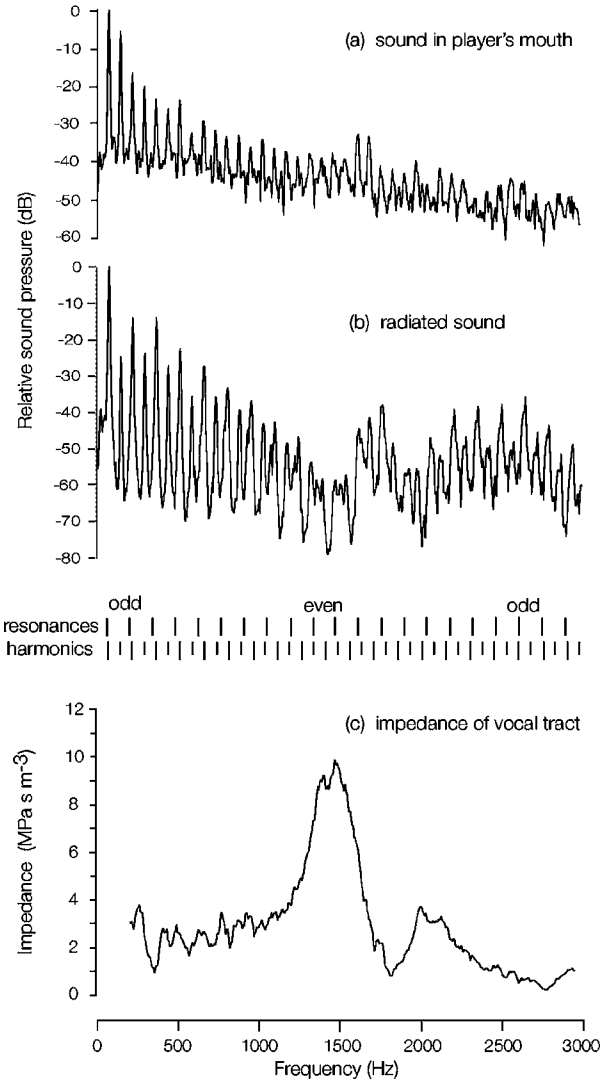


FIG. 6. A recording in the high-tongue configuration, but also illustrating the effects of resonance coincidence. The top graph is the sound spectrum measured inside the mouth, the middle is the sound spectrum outside the end of the yidaki at the same time, and the bottom is the impedance spectrum inside the mouth during playing. At low and high frequencies, the odd harmonics coincide with resonances of the pipe (vertical dashes). However, at around 1.5 kHz, even harmonics coincide with the resonances.

vocal tract (i.e., the airway between the lips and the glottis). Mukai (1989) reports that experienced wind players perform with the glottis almost closed. A nearly closed glottis produces relatively strong resonances at high frequencies because the coefficient of reflection is large. Consequently, at frequencies above several hundred Hz, the impedance seen at the lips is approximately that of an irregular tube closed at the glottis. The lungs, on the other hand, would produce a termination that is essentially resistive at high frequencies, and therefore an airway with open glottis exhibits rather weak resonances (data not shown). We propose that experienced yidaki players, like other wind players, also perform with the glottis nearly closed and that this is necessary to produce the relatively strong resonances that give rise to strong formants in the output sound, as discussed below. In a number of styles of yidaki playing, the glottis is used as an additional vibrating signal source, so players will presumably be used to keeping the glottis in a nearly closed configuration.

The resonances of the tract in playing are therefore somewhat analogous to those used to produce speech, the differences being that for speech the glottis is the vibrating signal source rather than the lips, and the lips are open. Vocal formants in the kHz range occur at frequencies which produce standing waves in the tract with pressure antinodes near the glottis and a flow antinode at the lip opening (Sundberg, 1977). In other words, they occur when the vocal tract is a most effective impedance matcher between the high impedance at the glottis and the low impedance of the radiation field outside the mouth. The formants radiated by the yidaki also occur when there is an impedance minimum at the lips and, we hypothesize, a pressure antinode near the glottis when it is nearly closed. Of course, in the sustained vowels of speech, the lips are at least somewhat open, whereas in yidaki playing they are almost shut. In speech, the mouth opening affects primarily the first vocal formant ($F1$, which occurs below about 1 kHz for all vowels). It is the second formant ($F2$) that is of interest here, because it falls approximately in the range 1 to 2 kHz. The frequency of the $F2$ in speech depends somewhat on mouth opening. Consequently, the strong yidaki formant could be expected to occur at frequencies comparable with, but not equal to, those of second speech formants for a similar tract configuration. Thus playing with a mouth configuration similar to that required to produce the vowel /i/, for example, will produce a yidaki sound whose formant frequency is similar to but not necessarily equal to that of the second formant in the vowel /i/. Further, to produce a yidaki formant, the mouth configuration must provide peaks in the tract impedance that have sufficiently high amplitude, so some vowel shapes with low tongue may not produce a clear formant, as is discussed below.

Figure 6 illustrates another effect that influences formants in the output sound. Again, the playing frequency is slightly above that of the first resonance—about 3 Hz in this case. Consequently, at a frequency approaching 1 kHz, the odd-even difference in the sound spectrum disappears, because in this frequency range the harmonics fall almost midway between resonances. At around 1.5 kHz, on the other hand, it is the even harmonics that benefit from the resonances of the pipe. This range is not far below the formant due to the vocal tract resonance. It might be possible, in principle, to observe broad and weak formants in the sound spectrum due only to this effect of an even or an odd harmonic happening to fall on a resonance of the yidaki: we call this the "harmonic coincidence" effect. For an experienced yidaki player, formants produced by this effect would be relatively small compared to those produced by the minima in the spectral envelope that coincide with peaks in Z_{tract}. Consequently, in this figure, as in Fig. 4, the maxima in the impedance of the vocal tract coincide with minima in the spectral envelope of the radiated sound. However, in Fig. 6 (but not Fig. 4), the formants at about 1.6 kHz is somewhat assisted by the near coincidence of a harmonic of the lip motion (here an *even* harmonic) with a resonance of the in-

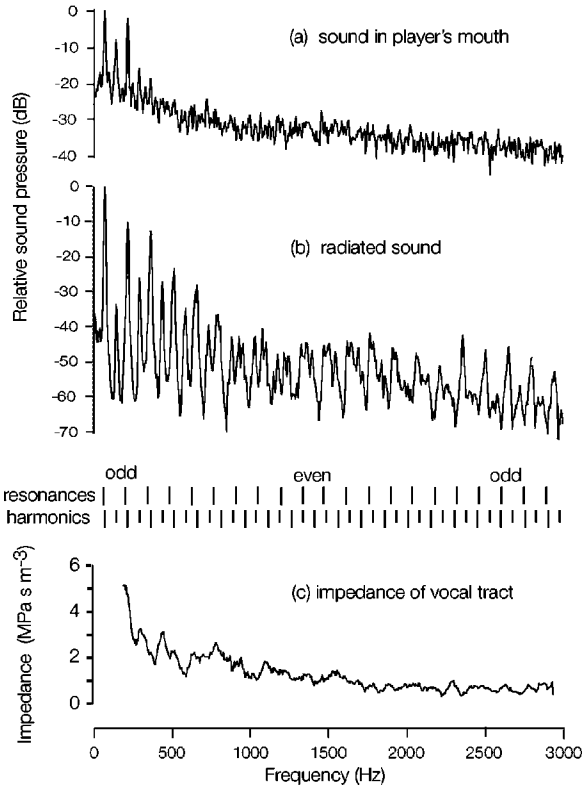


FIG. 7. The graphs are as for Figs. 4 and 6, but these data were taken during a note played with the low tongue configuration.

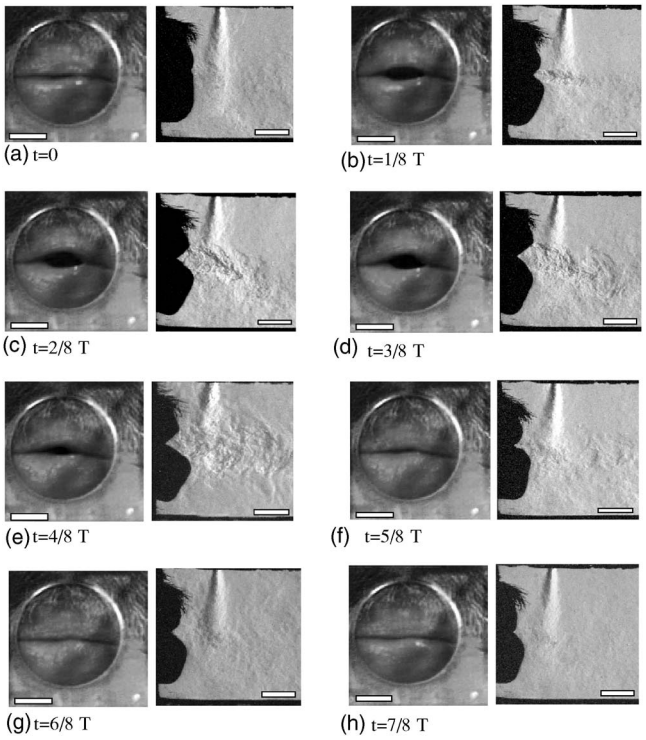


FIG. 8. The images on the left show one cycle of vibration of the player AT's lips, from the front, photographed with a high speed camera. The bar has length 10 mm. The side views (at right) were obtained stroboscopically by the schlieren flow visualization technique and are matched with the high speed images to show one complete cycle. The scale bars have length 10 mm.

strument. As the player can readily make small adjustments to the playing frequency, it is possible that the playing frequency is sometimes adjusted to take advantage of harmonic coincidence, though this was not investigated here.

Figure 7 shows the impedance measured inside the mouth and the sound produced for a configuration in which the player held the tongue low in the mouth. The peaks in Z_{tract} were not as large as for the high tongue configuration, and consequently there is none of the shaping of the spectral envelope of the sound as is produced with the high tongue figuration. The impedance of the vocal tract is so low over most of the frequency range that only the strong low harmonics of the lip motion produce components that are clearly visible above the turbulent noise present inside the mouth. There is no clear formant comparable to those near 1.7 kHz in Figs. 4 and 6. The coincidence effect does increase the odd harmonics around 2.5 kHz. In this example, and in others measured in the low tongue configuration, the odd harmonics in the signal recorded in the mouth were stronger than the neighboring even harmonics. Whether this is due to the shape of the lip opening (discussed below) or to sound transmission from the bore of the yidaki into the relatively low impedance load in the mouth, or to another cause, we do not know. However, one of its consequences might be the absence of a clear coincidence effect around 1.5 kHz.

C. Motion of the lips and air jet during playing

Images of the playing lips were produced in two different ways. In one, a high speed camera operating at 1000 frames per second filmed the lips from both the front and the side using the arrangement shown in Fig. 2. The side-view pictures thus taken are not shown here. Instead we present in side view the schlieren images, which show both the air jet and the lip position. These were produced stroboscopically using the arrangement shown in Fig. 3. Images from these two series are shown in Fig. 8. The images on the left of each pair are sequential pictures from one cycle. The pictures on the right, however, were each obtained in different cycles from the stroboscopic series and were matched to those from the high speed series by matching the shape of the lips at opening. The high speed film (both front and side view) is on our web site (Music Acoustics, 2005). The lip shapes shown here are qualitatively similar to the results reported by Wiggins (1988).

Figure 9 shows data from sets of images such as those on the left in Fig. 8, and the side view images taken simultaneously via the mirror (pictures not shown). The data are averaged over ten cycles from a series made with the high tongue configuration. The maximum camera speed (1000 frames per second) limits the time resolution, and hence limits the maximum frequency in an experimental spectrum $A(f)$ to 500 Hz. However, the shape of $A(t)$ is used as an input in a numerical model presented in our companion theory paper (Fletcher *et al.*, 2006). Yoshikawa (1995) reported the motion of the lips of horn players. In the low end of the range, their lips moved mainly along a horizontal axis (i.e., parallel to the bore of the mouthpiece), with a smaller opening motion in the vertical direction. Like that of the horn players in the low register, the lip motion reported here thus conforms

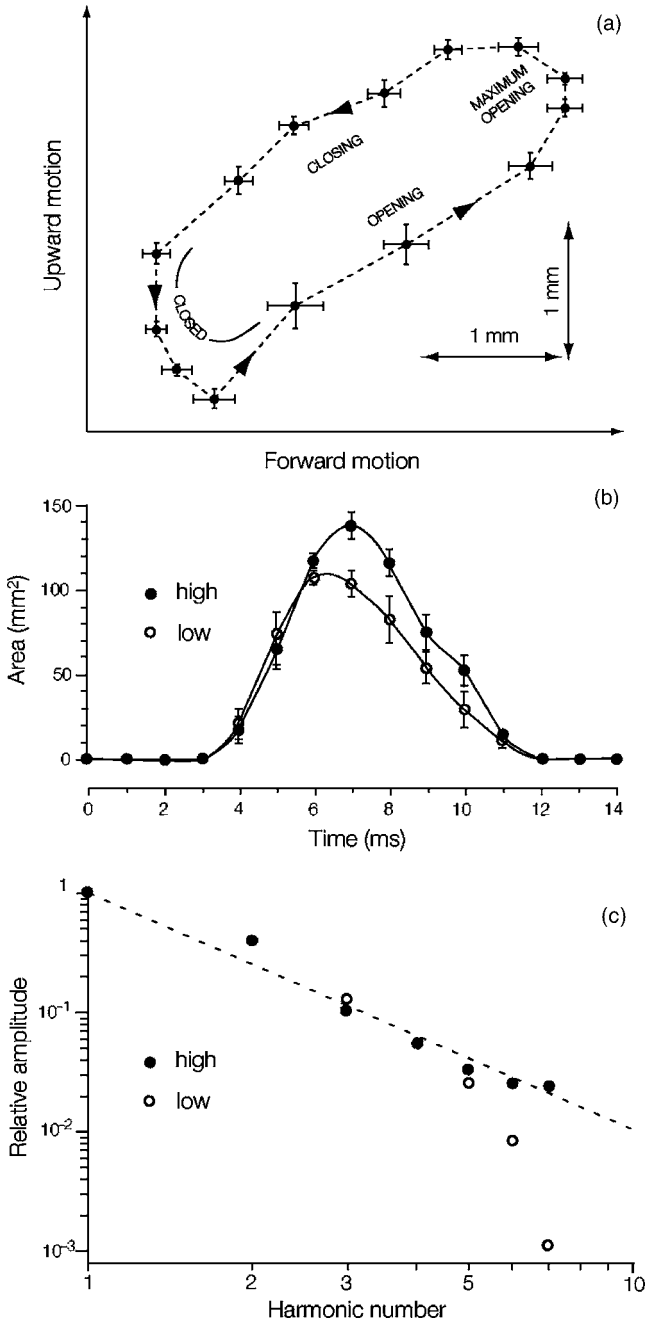


FIG. 9. Data from images such as those in Fig. 8, averaged over ten cycles. (a) shows how the average position of the upper lip varies throughout a complete cycle (bars show the standard error). (b) shows how the average area of the open space between the lips varies with time throughout a cycle for the high and low tongue configurations. (c) shows the relative amplitude of the harmonic components of the periodic variations in the lip area. The dashed line indicates the situation in which the harmonic components vary as n^{-2}.

to the "outward swinging door model" used to explain the operation of lip-valve instruments (Fletcher and Rossing, 1998) and the yidaki (Hollenberg, 2000).

The schlieren images in Fig. 8 show the motion of the jet emitted from the lips by mixing the jet with CO_2 from the "curtain" introduced as shown in Fig. 3. The density of carbon dioxide is about 1.5 times greater than that of air. However, the total volume of CO_2 is small so there is little reason to suspect that this alters significantly the behavior of the yidaki in general. Indeed, the period of the sound pressure oscillation (14.3 ms) remained substantially unchanged when the CO_2 was supplied to form the curtain. To reduce the effect of the CO_2 jets on the air jet, the flow of CO_2 was regulated so that the CO_2 curtain becomes almost invisible at the level of the opening of the lips. Some aspects of the motion of the jet are explained by the mouth opening. In Fig. 8(b), the lips have just opened and the jet is well-defined and narrow. In Figs. 8(b)–8(e) inclusive, the lips remain open and the jet grows broader. This is possibly the result of changing geometry of the jet separation from the lips as they open. It is interesting that the jet deviates noticeably downward from the axis of the yidaki [Figs. 8(e) and 8(f)]. The downward deviation may be caused by the changing geometry of the player's lips as they open or it may be due to a momentum transfer between the jet and the descending CO_2 stream. Once the lips close [Fig. 8(f)] the momentum of the jet keeps it moving and the disturbance of the "curtain" gradually disappears.

In a widely used Poiseuille-flow model for wind instruments, the relation between the flow of air into the instrument and pressure difference between the mouth and the instrument is calculated assuming conservation of energy between the mouth and the jet, followed by dissipation of kinetic energy in the jet when it mixes with the air in the instrument. When the air jet emerges from the lips it experiences viscous drag from the surrounding air, which leads to turbulent mixing because of the high speed of the jet. This can be clearly seen in Fig. 8. The important thing for sound generation in the instrument, however, is simply the input volume flow at a particular frequency multiplied by the instrument impedance at that frequency.

D. Idiomatic effects: Circular breathing and vocalization

Figure 10 shows aspects of the sound produced for another important playing configuration: the inhalation phase necessary for "circular breathing," during which the performer plays using the air in the inflated cheeks, with his mouth sealed from the lower vocal tract with the soft palate. This radical change in the vocal tract geometry makes a substantial change in the timbre and the necessity of frequent inhalation means that the change occurs often in continuous playing. Idiomatic playing makes a virtue of this necessity, so that the sound produced using the cheeks as reservoir is made an integral element of the rhythmic structure. Many of the rhythms used require simple alternation of inhalation and exhalation through the instrument. Typical mouth pressures vary from 0.8 to 2 kPa and flow rates from 0.1 to 0.3 l/s. The volume of air that can be expelled from the cheeks supports loud playing for only a fraction of a second, during which most players are unable to inspire deeply. Consequently, a simple alternation of inhalation and normal playing allows only shallow breathing.

For this reason, players will sometimes play for several seconds using the air in the lungs, and then refill the tidal volume during a series of three or four alternations between brief inhalations (while playing with the cheek reservoir) and

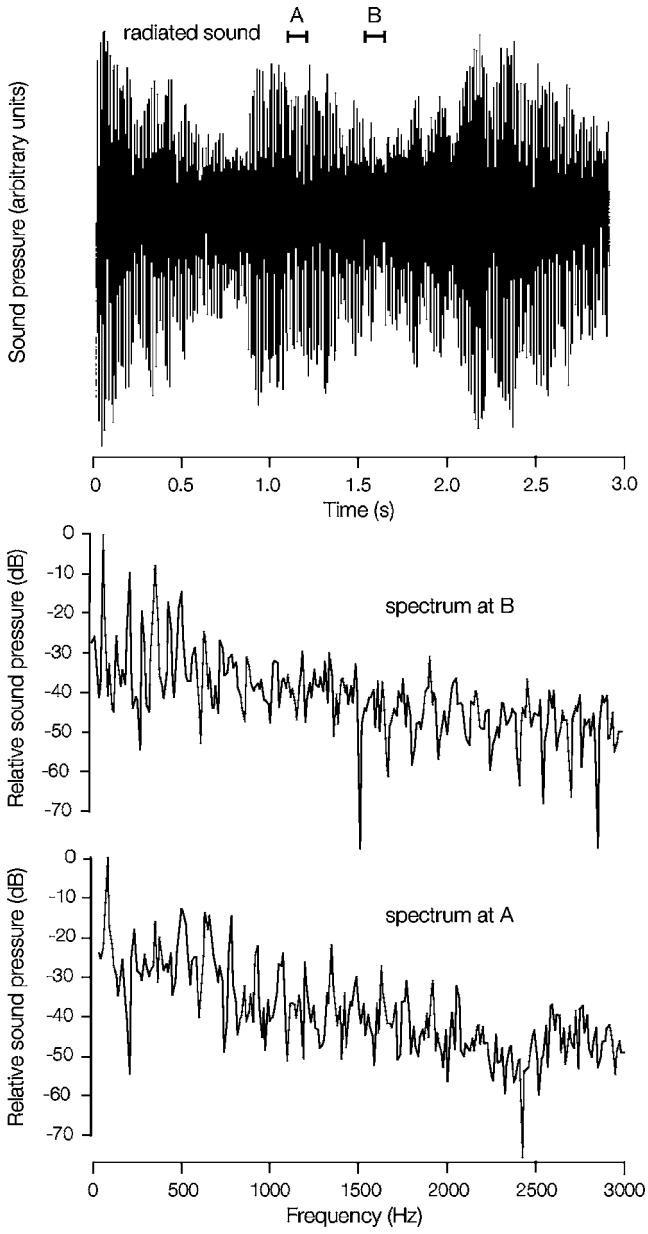


FIG. 10. Top: an oscillogram of a sound sample during which the player inhales through the nose three times, while continuing to play—an example of one of the common rhythms used in circular breathing. The amplitude falls during the inhalations and is largest during normal playing. Spectra of sounds during normal playing (a) and inhalation (b) are shown.

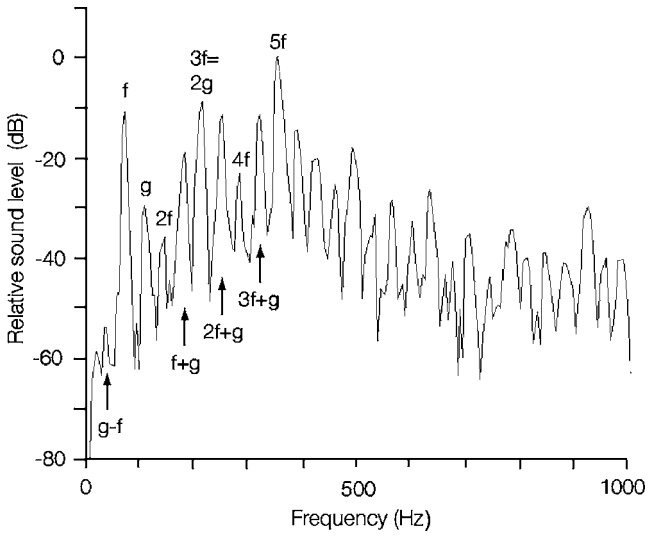


FIG. 11. The spectrum of a sound produced by playing a note with frequency $f=70$ Hz and simultaneously vocalizing at a frequency g, where $g=3f/2=105$ Hz.

normal playing. This may be used when preparing for or recovering from the use of a high flow rate, such as may be used to play one of the upper resonances.

An example of a regular series of inhalations is shown in Fig. 10. The waveform shows a series of three exhalations (large amplitude), each followed by a brief inhalation (small amplitude). The amplitude decreases during each of the inhalations, which may be explained by the application of the Young-Laplace relation to the muscles of the cheeks: it is easier to apply a higher pressure with a given muscular tension when the cheeks are highly curved. The finite mouth volume and the consequent brief period during inhalation (typically 0.2 s for this player) had the unfortunate consequence for this study that it proved impossible to make acceptable impedance measurements of the tract during this gesture.

The spectra in Fig. 10 contrast the spectra of the sound produced during one of the inhalations, during which the mouth alone was the resonator upstream from the lips and during the intervening exhalation.

A further series of different timbres can be obtained by vocalizing and playing at the same time. In this case, both the vocal folds and the lips act as pressure-controlled valves, operating at opposite ends of the vocal tract. Due to the strongly nonlinear interaction among lip motion, vocal fold motion, and the air flow, a range of heterodyne components are generated. Simply, if the vocal folds have opening area $S_1=\Sigma a(n)\sin(n\omega_1 t)$ and the lips have opening area $S_2=\Sigma b(n)\sin(n\omega_2 t)$, then the total flow, ignoring tract impedance effects, is proportional to S_1S_2 and thus has components at all frequencies $n\omega_1\pm m\omega_2$. An example is shown in Fig. 11. Player LH plays a steady note at frequency $f=70$ Hz, while simultaneously vocalizing at a frequency $g=3f/2=105$ Hz, a musical fifth above. The first several harmonics and heterodyne components are indicated in the figure. Because of the harmonic relation between the two frequencies, these components are equally spaced at frequencies $f/2=g/3=35$ Hz. Consequently, the pitch is heard at one octave below the lip fundamental, as the sound files demonstrate (Music Acoustics, 2005).

IV. CONCLUSIONS

When the player's tongue is raised close to the hard palate, the acoustical impedance spectrum of the vocal tract has maxima with values of 2–10 MPa s m^{-3} in the range 1 to 3 kHz—comparable with or larger than a typical impedance maximum of the yidaki in that frequency range. At the frequencies of these maxima, the spectral envelope of the radiated sound has minima, because the frequency components of acoustic current in the air jet entering the instrument

are reduced. In the frequency bands lying between the impedance maxima, the lower impedance of the tract produces a stronger acoustic current through the lips, resulting in a characteristic, strong formant in the radiated sound. These formants in the radiated sound correspond approximately to those of the second formant frequency produced in speech with the same tract configuration. The large variation between impedance maxima and minima, the variation that produces strong formants in the radiated sound, is consistent with a glottis nearly closed, the configuration used by experienced wind players studied by Mukai (1989, 1992). This suggests that learning to play with the glottis nearly closed may be important in developing technique on the yidaki. Broad and weak formants can also be observed when groups of even or odd harmonics coincide with a bore resonance.

The lips are open for about half of each cycle and operate approximately as "outward swinging doors," as described by Yoshikawa (1995) for the low range of the horn. The motion of the jet is relatively simple and supports the approximation of one-dimensional motion made in several analyses (Fletcher, 1993; Hollenberg, 2000). The present paper has been concerned entirely with experimental measurements. A theoretical analysis and justification of some of the conclusions will be presented in a companion paper (Fletcher *et al.*, 2006).

ACKNOWLEDGMENTS

We thank John Tann for technical assistance and the Australian Research Council for support. BL worked on this project during a vacation scholarship.

Amir, N. (**2004**). "Some insights into the acoustics of the didjeridu," Appl. Acoust. **65**, 1181–1196.

Amir, N., and Alon, Y. (**2001**). "A study of the didjeridu: normal modes and playing frequencies," Proc. International Symposium on Musical Acoustics, Perugia, edited by D. Bonsi, D. Gonzalez, and D. Stanzial, pp. 95–98.

Backus, J. (**1985**). "The effect of the player's vocal tract on woodwind instrument tone," J. Acoust. Soc. Am. **78**, 17–20.

Berio, L. (**1966**). *Sequenza V; Solo Trombone* (Universal, New York).

Brass, D., and Locke, A. (**1997**). "The effect of the evanescent wave upon acoustic measurements in the human ear canal," J. Acoust. Soc. Am. **101**, 2164–2175.

Caussé, R., Goepp, B., and Sluchin, B. (**2004**). "An investigation on 'tonal' and 'playability' qualities of eight didgeridoos, perceived by players," Proc. International Symposium on Musical Acoustics, Nara, Japan.

Clinch, P. G., Troup, G. J., and Harris, L. (**1982**). "The importance of vocal tract resonance in clarinet and saxophone performance—a preliminary account," Acustica **50**, 280–284.

Elliott, J. S., and Bowsher, J. M. (**1982**). "Regeneration in brass wind instruments," J. Sound Vib. **83**, 181–217.

Epps, J., Smith, J. R., and Wolfe, J. (**1997**). "A novel instrument to measure acoustic resonances of the vocal tract during speech," Meas. Sci. Technol. **8**, 1112–1121.

Erickson, R. (**1969**). *General speech: for trombone solo (with theatrical effects)* (Smith, Baltimore).

Fletcher, N., Hollenberg, L., Smith, J., and Wolfe, J. (**2001**). "The didjeridu and the vocal tract," Proc. International Symposium on Musical Acoustics, Perugia, edited by D. Bonsi, D. Gonzalez, and D. Stanzial, pp. 87–90.

Fletcher, N. H. (**1983**). "Acoustics of the Australian didjeridu," Australian Aboriginal Studies **1**, 28–37.

Fletcher, N. H. (**1993**). "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180.

Fletcher, N. H. (**1996**). "The didjeridu (didgeridoo)," Acoust. Aust. **24**, 11–15.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York).

Fletcher, N. H., Smith, J., Tarnopolsky, A., and Wolfe, J. (**2005**). "Acoustic impedance measurements—correction for probe geometry mismatch," J. Acoust. Soc. Am. **117**, 2889–2895.

Fletcher, N. H., Hollenberg, L. C. L., Smith, J., Tarnopolsky, A. Z., and Wolfe, J. (**2006**). "Vocal tract resonances and the sound of the Australian didjeridu (yidaki) II. Theory," J. Acoust. Soc. Am. **119**, 1205–1213.

Hollenberg, L. (**2000**). "The didjeridu: Lip motion and low frequency harmonic generation," Aust. J. Phys. **53**, 835–850.

Mukai, M. S. (**1992**). "Laryngeal movement while playing wind instruments," in Proc. International Symposium on Musical Acoustics, Tokyo, Japan, pp. 239–242.

Mukai, S. (**1989**). "Laryngeal movement during wind instrument play," J. Otolaryngol. Jpn. **92**, 260–270.

Music Acoustics (**2005**). http://www.phys.unsw.edu.au/~jw/yidakididjeridu.html

Smith, J. R., Henrich, N., and Wolfe, J. (**1997**). "The acoustic impedance of the Bœhm flute: standard and some non-standard fingerings," Proc. Inst. Acoust. **19**, 315–320.

Sundberg, J. (**1977**). "The acoustics of the singing voice," Sci. Am. **236**, 82–91.

Tarnopolsky, A., Fletcher, N., Hollenberg, L., Lange, B., Smith, J., and Wolfe, J. (**2005**). "The vocal tract and the sound of a didgeridoo," Nature (London) **436**, 39.

Tarnopolsky, A. Z., and Fletcher, N. H. (**2004**). "Schlieren flow visualisation technique: applications to musical instruments, especially the didjeridu," Proc. 18th International Congress in Acoustics, Kyoto, Japan, 4—9 April.

Tarnopolsky, A. Z., Lai, J. C. S., and Fletcher, N. H. (**2000**). "Flow structures generated by pressure-controlled self-oscillating reed valves," J. Sound Vib. **247**(2), 213–226.

Wiggins, G. C. (**1988**). "The physics of the didgeridoo," Phys. Bull. **39**, 266–267.

Wolfe, J., Smith, J., Tann, J., and Fletcher, N. H. (**2001**). "Acoustic impedance of classical and modern flutes," J. Sound Vib. **243**, 127–144.

Wolfe, J., Tarnopolsky, A. Z., Fletcher, N. H., Hollenberg, L. C. L., and Smith, J. (**2003**). "Some effects of the player's vocal tract and tongue on wind instrument sound." Proc. Stockholm Music Acoustics Conference (SMAC 03), edited by R. Bresin, Stockholm, Sweden, pp. 307–310.

Yoshikawa, S. (**1995**). "Acoustical behavior of brass player's lips," J. Acoust. Soc. Am. **97**, 1929–1939.

Vocal tract resonances and the sound of the Australian didjeridu (yidaki) II. Theory

N. H. Fletcher[a)]
Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

L. C. L. Hollenberg
School of Physics, University of Melbourne, Parkville 3010, Australia

J. Smith, A. Z. Tarnopolsky, and J. Wolfe
School of Physics, University of New South Wales, Sydney 2052, Australia

(Received 7 August 2005; accepted 8 November 2005)

The didjeridu (didgeridoo) or yidaki of the Australian Aboriginal people consists of the narrow trunk of a small Eucalypt tree that has been hollowed out by the action of termites, cut to a length of about 1.5 m, smoothed, and decorated. It is lip-blown like a trumpet and produces a simple drone in the frequency range 55 to 80 Hz. Interest arises from the fact that a skilled player can make a very wide variety of sounds with formants rather like those of human vowels, and can also produce additional complex sounds by adding vocalization. An outline is given of the way in which the whole system can be analyzed using the harmonic-balance technique, but a simpler approach with lip motion assumed shows easily that upper harmonics of the drone with frequencies lying close to impedance maxima of the vocal tract are suppressed, so that formant bands appear near impedance minima of the vocal tract. This agrees with experimental findings. Simultaneous vibration of the player's lips and vocal folds is shown to generate multiple sum and difference tones, and can be used to produce subharmonics of the drone. A brief discussion is given of player preference of particular bore profiles. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146090]

PACS number(s): 43.75.Fg, 43.70.Aj, 43.75.Rs, 43.72.Ct [DD] Pages: 1205–1213

I. INTRODUCTION

The Aboriginal peoples of Australia have lived in comparative isolation in this continent for something in excess of 40 000 years. They are divided into many nomadic tribes, each living in traditional lands and each speaking a different language but with many common customs. These people have developed several devices that have attracted wide attention, the best known being the boomerang, which returns to the hands of the thrower after quite a long circular flight, and the didjeridu, usually spelled "didgeridoo" in nonacademic literature. They have also developed a version of the "bullroarer" that is similar to that found in many other ancient cultures, and discovered how to imitate bird songs and other sounds using a "gum-leaf" from a Eucalypt tree pressed against the lips.[1]

Lip-blown wind instruments are common in many ancient cultures, their form depending upon availability of simple tube-like or conical structures to define the air column. Conch shells, for example, have been popular, as also have been tubes derived from the hollow stems of bamboo. The didjeridu was developed by the Yolngu tribes whose traditional lands are in what is now known as Arnhem Land, to the east of the modern city of Darwin on the central northern coastline of Australia. In the tribal language the instrument is called a yidaki or yiraki. In comparatively recent times its manufacture has spread to most Aboriginal craft communities and thence to many modern makers in other countries. Quite apart from its traditional use for the accompaniment of songs and ceremonies,[2,3] the didjeridu has also become widely used in "contemporary" music bands.[4]

The didjeridu is a simple wooden tube, typically about 1.5 m in length but sometimes made shorter, particularly in Central Australia. The tube is derived from the action of termites, which are common in Arnhem Land and which eat out the dry wood at the core of living Eucalypts, which is the major tree genus in Australia. Some species have quite narrow trunks, and the extent of termite damage can be assessed by knocking on the trunk with a heavy stick. When an adequately eaten trunk is found, the tree is cut down, traditionally using a sharpened stone, and a suitable section is cut from near its base. The termites leave a honeycomb-like galleried structure that can be cleaned out with a long stick and the inside of the tube rubbed to moderate smoothness, a process that is possible because of the growth rings in the wood. Some work with a sharp stone may be necessary at the two ends to give a smooth termination. The bark is stripped from the outside of the tube and it is then smoothed and usually painted with traditional tribal or family totem designs, using brown, black, and white pigments made from soil, charcoal, and minerals held together with plant gums. Finally, the smaller blowing end of the tube is smoothed for the player's lips by application of a ring of beeswax.

The sound of the didjeridu is a low-pitched drone with frequency usually about 70 Hz, maintained for minutes at a

[a)]Also at School of Physics, University of New South Wales, Sydney 2052, Australia. Electronic mail: neville.fletcher@anu.edu.au

time by the technique of "circular breathing" in which the cheeks are blown out to hold enough air to maintain sound production, the soft palate is momentarily closed, and there is a quick intake of breath through the nose. This technique is not unique to the didjeridu but has been used by traditional Indonesian flute players, and is now common among professional orchestral players of some woodwind instruments. The drone sound is normally modified by introducing emphasized frequency bands, known as formants by analogy with human vowel sound spectra, through change in the configuration of the upper vocal tract. These formants are usually introduced in a rhythmic manner to produce a musical pattern. Occasional accents are also introduced by momentary leaps to the second mode of the air column, at a frequency between about 2.5 and 3 times the drone frequency, depending upon the shape of the instrument bore. Finally, while maintaining the lip vibration, the player may also occasionally vibrate his vocal folds as in singing to produce additional sounds or even subharmonics of the drone. These techniques are often used to imitate the sounds made by native animals or to illustrate the text of a song or story.

The acoustics of the didjeridu tube is essentially trivial, since it consists of a tube of defined though irregular geometry with enhanced wall losses because of roughness and porosity. Details of these parameters determine the distinction between an instrument considered "good" by the player and one regarded as "poor,"[5–7] as will be discussed later. Interest however centers upon the coupling of the vocal tract to the instrument and on the vibration of the player's lips. There has been rather little investigation of these problems, except for work by Fletcher,[8,9] Wiggins,[10] Hollenberg,[11] and members of the present project team.[12]

The purpose of the present paper is to give a formal analysis of the acoustics of the played didjeridu, including the important contributions of the player's lungs and vocal tract. Some consideration will also be given to the effect of vocalizing while playing. Only brief mention will be given here of experimental studies by our project team in order to provide a comparison with calculated results. Detailed discussion of the measurements and experiments conducted is given in a companion paper,[13] and a brief discussion of related experimental results has recently been published.[14]

II. ANALOG NETWORK MODEL

Analysis of the acoustic behavior of the entire system consisting of lungs, glottis, vocal tract, lips, and didjeridu tube is most conveniently carried out using an electrical network analog,[15–17] in which voltage represents acoustic pressure and current represents acoustic volume flow. Since the whole system is very complicated, it is helpful to treat it in sections, beginning with the lungs.

A. Lung impedance

The lung is a complex quasifractal structure of branching tubules.[18–20] In a simplified version, starting just below the vocal folds the main tube branches into two bronchi which continue to branch into pairs of tubules through 16 orders. After 16 orders of branching the tubules are terminated by about 65 000 small alveolar sacs which contribute most of the stored air volume. There are, of course, several spatial constraints to be obeyed by this structure so that the tubules and finally the alveolar sacs can fit within the overall volume occupied by the lungs in the body. The whole structure is somewhat elastic, and the pressure inside it is maintained by muscle tension on walls of quite large total mass.

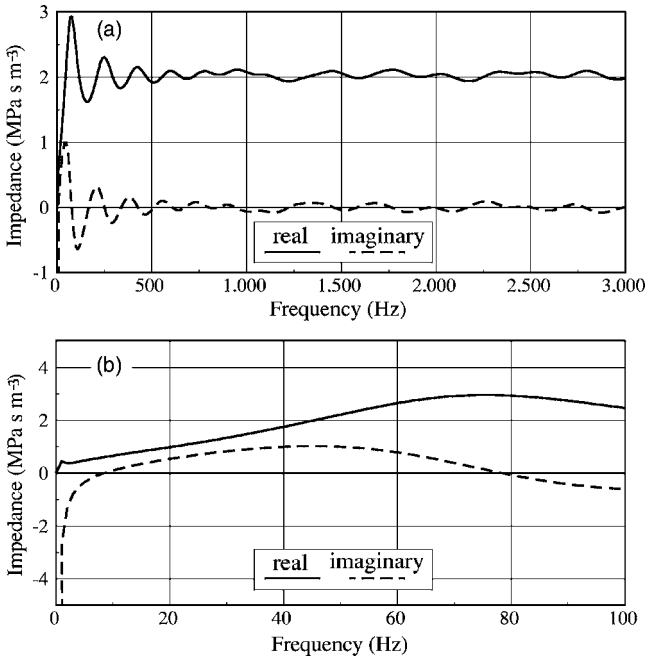


FIG. 1. Calculated input impedance just below the vocal folds for the idealized lung model (a) over an extended frequency range, and (b) over the range of interest for the lip vibrations. The characteristic impedance of the bronchus is about 2 MPa s m^{-3}.

As a simple model for the lungs that is a good approximation to reality, we take the tubule diameter ratio after branching to be 0.8, rather than the idealized value of $2^{-1/2}$, and take the tubule length ratio also to be 0.8. This allows the geometrical packing constraints to be satisfied and gives a total tubule volume of about 1 liter, leaving the remaining 3 liters of lung capacity to be provided by the alveolar sacs. It is straightforward to calculate the acoustic input impedance for this idealized model, and the results are shown in Fig. 1. Over the range of interest for the vibration of the lips, near 70 Hz, the impedance is almost purely resistive and about equal to the characteristic impedance of the bronchus, 2×10^6 Pa s m^{-3}, and the same is true in the range from about 0.5 to 3 kHz that is of interest for the vocal-tract resonances leading to sound formants.

As a consequence of this analysis, it is a good approximation to represent the lungs in the network model as shown in Fig. 2(a). Muscular pressure on the lungs is represented by a dc voltage source in series with a very large inductance representing the mass load (inertance) of the associated muscular tissue. The acoustic compliance of the air volume in the alveolar sacs is represented by a simple capacitance, while the impedance of the network of tubules contributes a simple resistance with magnitude about equal to the characteristic impedance of an infinite tube of diameter equal to that of the subglottal bronchus. Such a model reproduces the behavior shown in Fig. 1.

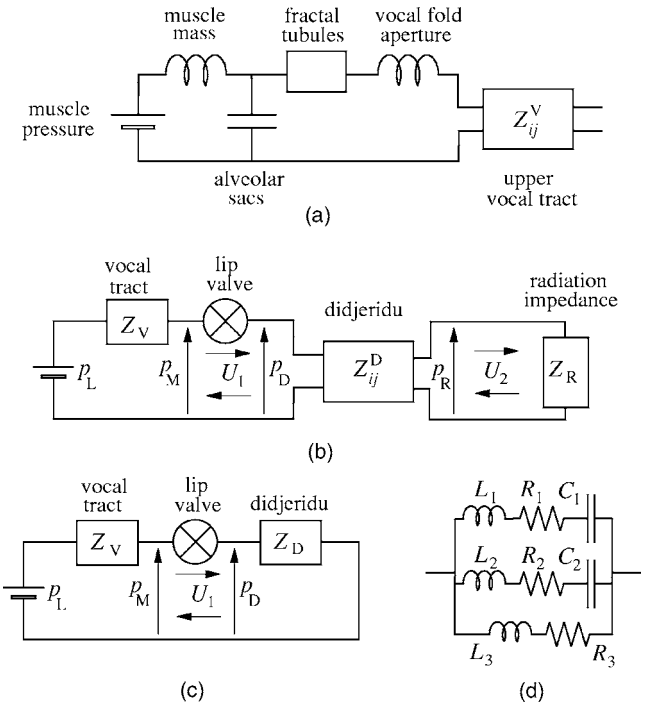


FIG. 2. (a) Electrical network representing the lungs, glottis, and vocal tract of the didjeridu player. (b) Simplified network representing the entire didjeridu+player system. (c) Further simplified network used for calculation. (d) Simple model used for the vocal tract impedance Z_V.

Above the junction of the two bronchi, we find the constriction of the vocal folds. Since these folds are inactive during normal playing, the larynx constriction can be represented in the model by a simple series inductance, as shown. It appears that skilled players on most wind instruments partly close (adduct) their vocal folds so that the area of the remaining aperture is small, a feature first noted and examined in detail by Mukai.[21,22] This provides a significant series inertance between the upper vocal tract and the lungs and serves to isolate the upper-tract resonances to a large extent from the damping influence of the resistive lung impedance. For typically adducted vocal folds, the opening is about 2×10 mm and the effective fold thickness, including end corrections, perhaps 5 mm, giving an acoustic inertance at frequency ω of about 300ω Pa s m^{-3}, which is about 2–6 MPa s m^{-3} in the frequency range 1–3 kHz that is of interest for formants. This must be compared with characteristic impedance $\rho c/S \approx 1$ MPa s m^{-3} of the trachea and the estimated subglottal lung impedance, illustrated in Fig. 1, which is nearly purely resistive and about 2 MPa s m^{-3} in magnitude. It is therefore evident that the vocal folds can have a significant effect on the resonances of the vocal tract provided they are well adducted.

B. Vocal-tract impedance

Above the vocal folds, which have already been discussed, the trachea is a nearly uniform tube of diameter about 20 mm until it reaches the glottis, above which the geometry is complicated by the movable lower jaw and tongue. Magnetic resonance imaging (MRI) studies[12] of one of us (LH) while playing on a modified didjeridu provide some numerical values for the geometrical parameters for both the "uninflected drone" (low tongue) and "high formant" (high tongue) configurations. As will be seen from the analysis to follow, the acoustical quantity of importance in determining the spectral envelope of the didjeridu sound is the frequency dependence of the input impedance Z_V of the vocal tract at the lips, as shown in Fig. 2(b). This quantity, measured by a modification of an impedance head developed by our group,[23,24] is discussed in detail in the companion experimental paper.[13] In the low-tongue position, the mouth simply extends the trachea into a uniform tube up to the lips, giving a rather low value for Z_V. In the high-tongue configuration, normally used to produce sounds with pronounced formants, the tongue is raised along its whole length and brought close to the teeth, so that it forms a narrow passage of effective diameter about 10 mm from the root of the tongue up to the lips. By changing the position of the lower jaw and tongue, it is also possible to convert this passage from almost cylindrical to nearly conical, with obvious influence upon the tract resonances, while the narrow duct diameter near the lips greatly increases the value of Z_V. This effect has been briefly described, both for the didjeridu and for other wind instruments, in a recent conference publication.[25]

A detailed model for the impedance Z_V of the whole vocal tract at the lips can be built up using the analog network shown in Fig. 2(a), the variations in the geometry of the upper vocal tract being detailed through the coefficients $Z_{ij}^V(\omega)$. In the analysis to follow, however, a very simplified model will be used for the vocal-tract impedance, as shown in Fig. 2(d). This consists of two series-resonant LRC circuits in parallel, representing the resonances of the upper vocal tract, the characteristic impedance $(L/C)^{1/2}$ of these circuits being chosen to match the expected impedance of the vocal tract for the high-tongue configuration, and with a further parallel LR branch representing the impedance of the glottal constriction and the lungs.

C. Didjeridu impedance

The acoustic properties of the didjeridu can be described by four impedance coefficients $Z_{ij}^D(\omega)$, the values of which can be calculated from the measured dimensions of the tube and the estimated roughness of the walls. Most genuine didjeridu tubes flare progressively along their length so that their shape is roughly that of a truncated cone. For such a tube, ideally open at end 2, the impedance maxima at the mouth occur at frequencies ω_n such that

$$k_n L' = n\pi - \arctan\left(\frac{k_n d_1 L'}{d_2 - d_1}\right), \tag{1}$$

where $k_n = \omega_n/c$, c is the speed of sound in air, d_1 is the tube diameter at the blowing end, d_2 is the larger diameter at the open end, L is the tube length, and $L' \approx L + 0.3d_2$ is the effective tube length including the open-end correction.[9] For a cylindrical tube with $d_2 = d_1$, the mode frequencies ω_n given by (1) follow a sequence 1,3,5,… times $\pi c/2L'$, while for a complete cone of the same length the sequence is 2,4,6,… times $\pi c/2L'$. For an intermediate geometry of small flare, the mode frequencies are given by

$$\omega_n \approx \left(n - \frac{1}{2}\right)\frac{\pi c}{2L'}\left\{1 + \left[1 + \frac{4(d_2 - d_1)}{\pi^2 d_1 (n - (1/2))^2}\right]^{1/2}\right\}. \quad (2)$$

These frequencies lie between those of the limiting sequences mentioned above for cylindrical and for conical tubes of the same length, and thus might be described as a "stretched quasiharmonic sequence." A real didjeridu tube, being a natural object hollowed out from a tree trunk by termites, does not have quite this regular form, though the behavior of its lower modes is generally well approximated by the flared-tube relation (2).

Experimental values for the impedance coefficients $Z_{ij}^D(\omega)$ can be obtained by measuring the input impedance of the tube from each end with the opposite end either open or closed, in separate measurements, or can be calculated if the geometry is assumed to be regular. If the shape of the bore of the didjeridu is assumed to be a uniform cylinder of length L, as it may be in some laboratory experiments using plastic pipe, then the impedance coefficients have the simple forms

$$\begin{aligned} Z_{11}^D &= Z_{22}^D = -jZ_0 \cot kL \\ Z_{12}^D &= Z_{21}^D = -jZ_0 \csc kL, \end{aligned} \quad (3)$$

where $Z_0 = 4\rho c/\pi d^2$ is the characteristic impedance of the tube, assumed to have diameter d, and $k \approx \omega/c - j\alpha$. The quantity α allows for viscous and thermal losses to the tube walls according to the relation

$$\alpha \approx 10^{-5}\beta\frac{\omega^{1/2}}{d}, \quad (4)$$

in SI units. The numerical parameter β is about 2.4 for perfectly smooth walls, but may become much larger than this for the rough walls characteristic of a naturally produced didjeridu.

If, as is typical of genuine instruments, the bore has nearly the shape of a flaring truncated cone with diameter d_1 at the blowing end and diameter d_2 at the open end, then the impedance coefficients have the more complicated forms[26,27]

$$\begin{aligned} Z_{11}^D &= -\frac{4j\rho c}{\pi d_1^2}\left[\frac{\sin(kL + \theta_2)\sin\theta_1}{\sin(kL + \theta_2 - \theta_1)}\right] \\ Z_{22}^D &= +\frac{4j\rho c}{\pi d_2^2}\left[\frac{\sin(kL - \theta_1)\sin\theta_2}{\sin(kL + \theta_2 - \theta_1)}\right] \\ Z_{12}^D &= Z_{21}^D = -\frac{4j\rho c}{\pi d_1 d_2}\left[\frac{\sin\theta_1 \sin\theta_2}{\sin(kL + \theta_2 - \theta_1)}\right], \end{aligned} \quad (5)$$

where $\theta_1 = \arctan kx_1$ and $\theta_2 = \arctan kx_2$, with x_1 and x_2 being the axial distances of the ends 1 and 2 from the apex of the cone. Related expressions can also be derived for the input impedance Z_D when the bore is terminated by a given impedance,[27] in the present case the radiation impedance Z_R. These coefficients Z_{ij}^D must of course include the real parts due to viscous and thermal losses at the tube walls, as detailed in (4). If d_2 is the diameter and A_2 the area of the open end, then the radiation impedance is also a complex quantity which has the approximate form[17]

$$\begin{aligned} Z_R &\approx 7 \times 10^{-4}\omega^2 + 0.5j\frac{\omega}{d_2} \quad \text{if } kd_2 \ll 1 \\ &\approx \frac{\rho c}{A_2} \approx \frac{500}{d_2^2} \quad \text{if } kd_2 \gg 1 \end{aligned} \quad (6)$$

in SI units.

III. LIP-VALVE OPERATION

The lip valve is in reality a complex mechanism with many degrees of freedom. The player's lips are somewhat soft and flexible, their geometry is complicated, and they usually close completely for part of each vibration cycle, thus introducing very significant nonlinearity. As has been found with vocal-fold models, construction of any detailed model of the vibrating lips is a very major undertaking, and the appropriate model may well differ greatly from one player to another, or indeed during different types of performance by a single player. For these reasons, among others, it is preferable to develop a model that is as simple as possible, while still incorporating the major features of a real lip valve.

Experimental studies of players of brass-instruments players[28–31] have shown that the predominant lip motion at low frequencies can be described as that of "outward-swinging doors," or the configuration $(+, -)$ in the notation of Fletcher,[32] where the first symbol indicates the action on the valve of a steady pressure on the upstream side and the second symbol the corresponding action of a downstream pressure. For fundamentals of high frequency, these studies of trumpet and trombone players show that a better approximation is that of "sliding doors" with configuration $(+, +)$. Other studies have described the motion in terms of a combination of both types of motion[29] and some as surface waves propagating on the lips. To maintain generality we shall initially make no assumption about which of these configurations is a more appropriate approximation for the didjeridu and write the configuration symbol as (σ_1, σ_2) where, in the present case, $\sigma_1 = +1$ and σ_2 may be either $+1$ or -1 depending upon the lip model adopted.

Our own experimental studies of didjeridu players, described in detail in the companion paper,[13] show that the lips are open for only about half of an oscillation cycle, that the lip opening is typically nearly elliptical, and that both elliptic diameters expand and contract to some extent together. The area S of the lip opening thus has approximately the form

$$S(t) = A\langle\cos\omega t - \cos\zeta\rangle^{\mu}, \quad (7)$$

where A is a constant, $1 \leqslant \mu \leqslant 2$, and the notation $\langle\ldots\rangle$ is intended to imply that the enclosed expression is set to zero if it is negative. Here, ζ/π is the fraction of the period for which the lip valve is open, and since this is typically about half of the period for didjeridu players, $\zeta \approx \pi/2$, though this may vary somewhat with playing style. The expression (7) is usually referred to as a "transparent closure" model of the lip vibration, for obvious reasons.

While the opening and closing of the lips may not be exactly symmetrical in practice because of hysteresis effects in lip contact, we neglect this complication and write

$$S(t) = \sum_{0}^{\infty} s_n \cos n\omega t. \tag{8}$$

Consideration of (7) for the transparent closure model indicates that we should expect that $s_n \propto n^{-1}$, but this needs to be checked by experiment. Stroboscopic measurements on the lips of a didjeridu player, as shown in Fig. 8 of our experiment paper,[13] allow the opening area to be plotted as a function of time, and from this the coefficients s_n can be determined. To a reasonable approximation, as shown in Fig. 9 of that paper, it is found that $s_n = n^{-2}s_0$, which differs from the predictions of the transparent closure model. The difference can be ascribed to a "soft" closure of the lips instead of the "sharp" closure when they pass through each other in this unrealistic model. The experimental result will be used in the calculations to follow, though it is possible that the player may be able to modify this relation to some extent by changing the tension in the lip muscles. Indeed, Fig. 9(c) of the experiment paper shows just such an effect in the "low-tongue" position, with the amplitude of higher harmonics s_n for $n > 5$ falling in amplitude even more rapidly than n^{-2}.

Since it is not the purpose of this paper to calculate lip motion in any detail, an outline based upon a transparent closure model referred to above in which, instead of a realistic nonelastic collision between the surfaces of the lips, it is assumed that each lip executes a sinusoidal motion and effectively passes through the other lip.[11] While this is clearly unrealistic, the properties of the collision can be adequately modeled by ascribing an appropriate damping to the motion of each lip. The outline of this model given below expresses it in one-dimensional form for simplicity.

Suppose that the upstream pressure is p_M and the downstream pressure p_D, as shown in Fig. 2(b), and that both act upon an effective lip area A. Let the lips be identical and have combined effective mass m and damping γ. If z measures the linear lip opening, then an adequate approximation to the lip behavior on this model is

$$m\frac{d^2z}{dt^2} + \gamma\frac{dz}{dt} + m\omega_0^2(z - z_0) = A(\sigma_1 p_M + \sigma_2 p_D), \tag{9}$$

where ω_0 is the natural frequency of the lip vibration as determined by their mass and tension. In the swinging-door case, the parameters m and γ should be defined in terms of moments of inertia and centroid of pressure, but this makes no basic difference.

Discussion of a simplified version of this problem, in which the equilibrium opening is great enough that the valve does not close and the aperture width W is constant rather than varying in synchronism with the aperture height, has been given elsewhere.[32,33] The conclusion is that the drone frequency must be a little above the didjeridu resonance in the case of a $(+, -)$ lip valve and a little below the resonance in the case of a $(+, +)$ valve. The player is presumably able to adjust the lip motion between these two configurations, as can trumpet players, though the total range of variation in the case of the didjeridu is only about a semitone (6%). Informal observations of competent didjeridu players show that they typically raise the drone frequency when wishing to emphasize formants.

A general approach that properly includes all nonlinearities and treats finite vibration amplitudes is based upon the "method of slowly varying parameters" described by Bogoliubov and Mitropolsky.[34] In this approach it is assumed that the lip oscillation can be described by the expression

$$z = z_0 + a\sin(\omega_0 t + \phi), \tag{10}$$

and that both the amplitude a and the phase ϕ are slowly varying functions of time in the sense that they change by only a small amount over one period $2\pi/\omega_0$ of the oscillation. If (9) is rewritten in the form

$$\frac{d^2z}{dt^2} + \omega_0^2(z - z_0) = g\left(z, \frac{dz}{dt}, p_M, p_D\right), \tag{11}$$

then it can be shown[34,35] that

$$\frac{da}{dt} \approx \frac{1}{\omega_0}[[g\cos(\omega_0 t + \phi)]] \tag{12}$$

$$\frac{d\phi}{dt} \approx -\frac{1}{a\omega_0}[[g\sin(\omega_0 t + \phi)]], \tag{13}$$

where the notation $[[\ldots]]$ is to be taken to imply that only terms varying slowly in time relative to $\sin\omega_0 t$ are to be retained. When the system settles into a steady state, the vibration amplitude a is constant and so too is $d\phi/dt$, implying that the new steady vibration frequency is $\omega_0 + d\phi/dt$. The qualitative results are the same as outlined above for the simple linearized model.

While detailed consideration of lip vibration is clearly necessary in a complete model for didjeridu sound production, creation of such a model is not the purpose of the present paper. Indeed, it will be sufficient to note that a model for lip vibration is possible and to specify its predictions by assuming reasonable values for the vibration amplitude and frequency. The frequency, as shown by (13), depends upon lip tension through the value of ω_0, upon the didjeridu and vocal-tract impedances, and upon the nature of the lip motion as encoded in the parameters σ_2 and μ.

IV. SYSTEM ANALYSIS

In this section two approaches will be explored. The first, which is potentially able to predict the operation of the whole system in detail, relies upon the "harmonic-balance" approach to the behavior of the whole system. Unfortunately this approach is quite complex and necessarily numerical, so that the general trend of its conclusions is not readily apparent. The second approach reduces the whole system to the simplest possible approximation by considering only first-order terms. While the conclusions derived are certainly not quantitatively accurate, this approach does show clearly the general trends and, in particular, the way in which the spectrum of the output sound depends upon the vocal-tract parameters, which is the main purpose of the whole analysis.

In both cases the system to be modeled is that shown in Fig. 2. Part (a) of this figure gives a model for the player's

lungs and vocal tract, while part (b) compresses this into a simpler system and includes the didjeridu as well. Finally, part (c) compresses the didjeridu subsystem further to give the electrical analog network that will be analyzed. The essential feature of this progressive simplification is that both the vocal tract and the didjeridu itself are ultimately represented by simple input impedances Z_V and Z_D, respectively, on the two sides of the lip valve. The values of these two quantities can be readily calculated, as has already been discussed.

A. Harmonic-balance approach

In the harmonic-balance approach[36] the procedure is as follows, after the relevant values of $Z_V(\omega)$ and $Z_D(\omega)$ have been evaluated. (i) Assume reasonable values for mouth pressure p_M, lip vibration frequency ω_0, and lip vibration amplitude a. Take the pressure p_D at the input to the didjeridu tube to be zero. (ii) Calculate the volume flow through the lips as a function of time, using the Bernoulli equation. (iii) Take the Fourier transform of this flow and use it to calculate the Fourier components of the pressures p_M and p_D in the mouth and didjeridu input. (iv) Convert these pressures back into the time domain, or perhaps simply deal with the component at the fundamental frequency, which nearly matches the vibration frequency of the lips. (v) Use this driving pressure difference to recalculate the frequency and amplitude of the lip vibration, using either the "transparent closure" model, or preferably a more realistic model, for lip motion and the solution techniques detailed in Sec. III. (vi) Iterate steps (ii) through (v) until convergence is achieved. (vii) Calculate the resulting sound power radiation from the open end of the didjeridu tube.

It is clear that this is a rather complex though straightforward computational process, so that no simple general conclusions can be drawn except by examining behavior in many particular cases. For this reason such a detailed computation will not be carried out or discussed further at this time.

B. First-order approximation

Referring to Fig. 2(c), and denoting the lung pressure by p_L, the mouth pressure by p_M, and the input pressure to the didjeridu by p_D, the following equations can be written:

$$p_L - p_M = Z_V U, \tag{14}$$

$$p_D = Z_D U, \tag{15}$$

$$U = \left[\frac{2(p_M - p_D)}{\rho}\right]^{1/2} S(t), \tag{16}$$

where Z_V is the input impedance of the vocal tract at a position just inside the lips and Z_D is the input impedance of the didjeridu just outside the lips, $S(t)$ is the opening area of the lips, and U is the acoustic volume flow through the lips. Substituting (14) and (15) in (16) then gives

$$U = \left(\frac{2}{\rho}\right)^{1/2} [p_L - (Z_V + Z_D)U]^{1/2} S(t)$$
$$= \left(\frac{2}{\rho}\right)^{1/2} [p_L - ZU]^{1/2} S(t), \tag{17}$$

where we have written $Z \equiv Z_V + Z_D$ for convenience. Suppose now that

$$U = \sum_0^\infty (u_n \cos n\omega t + v_n \sin n\omega t). \tag{18}$$

Recalling the Fourier expansion (8) for $S(t)$ and writing p_0 for the steady component of the mouth pressure p_M as given by (14),

$$\sum_{n=0}^{\infty} (u_n \cos n\omega t + v_n \sin n\omega t)$$
$$= \left(\frac{2}{\rho}\right)^{1/2} \left[p_0 - \sum_{n=1}^{\infty} Z_n (u_n \cos n\omega t + v_n \sin n\omega t)\right]^{1/2}$$
$$\times \sum_{m=0}^{\infty} s_m \cos m\omega t, \tag{19}$$

where Z_n is the value of $Z_V + Z_D$ at frequency $n\omega$. Note that the introduction of p_0 has removed the zeroth-order term from the summation inside the parentheses because the didjeridu tube impedance $Z_D = 0$ when $\omega = 0$. In (19), since $Z_n = R_n + jX_n$ is a complex quantity, the symbol j must be interpreted as an operator meaning "advance the phase by $\pi/2$."

As a reasonable approximation in the case of the didjeridu, we might assume that $p_0 \gg Z_n u_n, Z_n v_n$ for all n, so that the factor $[p_0 - \Sigma]^{1/2}$ on the right in (19) can be simply approximated by $p_0^{1/2}[1 - (1/2p_0)\Sigma]$, where in each case Σ represents the summation in this factor in (19). Since typically $s_1 \approx s_0 \gg s_n$ for all $n > 1$, we can also treat s_0, s_1, u_0, u_1, and v_1 as being all of zeroth order, while $v_0 = 0$ and higher terms are all of first order or less. Even with this simplification, however, and limiting consideration to harmonics up to N, the result is an algebraically complicated, though sparse, $N \times N$ matrix equation containing nonzero elements M_{ij} for $j = i$ and $j = i \pm 1$.

While this matrix could be solved for a specific case, for a semiquantitative analysis it is appropriate to simplify (19) still further by assuming that $s_n \ll s_0$ for all $n > 0$, rather than just for $n > 1$. This carries with it the additional result that $u_n, v_n \ll u_0$ for all n. Physically this amounts to the assumption that, instead of closing in each cycle, the lips remain substantially open but vibrate in a nonlinear manner so as to generate all the terms s_n with magnitudes approximately proportional to $1/n^2$. Selection of terms in $\cos n\omega t$ and $\sin n\omega t$ in Eq. (19) then leads to the simpler results

$$u_n[(2p_0\rho)^{1/2} + s_0 R_n] + s_0 X_n v_n = 2p_0 s_n$$
$$v_n[(2p_0\rho)^{1/2} + s_0 R_n] = s_0 X_n u_n, \tag{20}$$

which can be combined to give

$$u_n = \frac{2p_0 s_n[(2p_0\rho)^{1/2} + s_0 R_n]}{[(2p_0\rho)^{1/2} + s_0 R_n]^2 + s_0^2 X_n^2} \tag{21}$$

$$v_n = \frac{2p_0 s_0 s_n X_n}{[(2p_0\rho)^{1/2} + s_0 R_n]^2 + s_0^2 X_n^2}. \tag{22}$$

These two equations can be further combined to give the flow magnitude at frequency $n\omega$ as

$$(u_n^2 + v_n^2)^{1/2} = \frac{2p_0 s_n}{[(2p_0\rho)^{1/2} + s_0 R_n]^2 + s_0^2 X_n^2]^{1/2}}, \tag{23}$$

which can be simply evaluated explicitly.

Since the quantity of interest is not the acoustic flow into the didjeridu from the lips, but rather the acoustic radiation from the open end of the tube, it is necessary to calculate the transfer function involved. Referring to the network analog shown in Fig. 2(b) and using a new notation with $U_1(n) \equiv u_n + jv_n$ and $U_2(n)$ being the acoustic flow from the tube into the radiation impedance Z_R, we can write

$$p_1(n) = Z_{11}^D U_1(n) - Z_{12}^D U_2(n) \tag{24}$$

$$p_2(n) = Z_{21}^D U_1(n) - Z_{22}^D U_2(n). \tag{25}$$

The minus signs in these two equations arise from the fact that both currents U_1 and U_2 have been taken to flow clockwise, as in Fig. 2(b), rather than symmetrically with respect to Z_{ij} as is sometimes the convention. At the open end of the tube $p_R = Z_R U_2$, so that (25) shows that

$$U_2(n) \approx \frac{Z_{21}^D}{Z_{22}^D + Z_R} U_1(n), \tag{26}$$

and this is the transfer function that is sought. Since normally $Z_R \ll Z_{22}^D$, with most of the resistive losses due to wall effects in the tube and the reactive term easily included as a simple end correction, this transfer function is approximately equal to Z_{21}^D / Z_{22}^D, which has maxima at the same frequencies as does the input impedance at the lip end of the tube.

The radiated acoustic power at frequency $n\omega$ is therefore

$$\Pi(n) = R_R(n\omega) U_2(n)^2 = R_R(n\omega) \left[\frac{Z_{21}^D}{Z_{22}^D + Z_R} U_1(n) \right]^2, \tag{27}$$

where R_R is the resistive component of the radiation impedance Z_R as given by (6) and $U_1(n) \equiv u_n$ in the previous analysis.

V. PREDICTIONS OF THE MODEL

The implications of this analysis can now be examined. The simplest case is that in which the didjeridu is a narrow uniform cylinder, for the impedance coefficients Z_{ij}^D for the tube are then easily expressed, as in (3), and the radiation resistance is simply proportional to the square of the frequency, as given by the lower range of (6). The expression for the radiated sound power, as given by (27), can then be evaluated explicitly once the relation of the drone frequency to the first resonance of the tube has been specified.

TABLE I. Model parameter values (high tongue).

Didjeridu tube length	L	1.3 m
Didjeridu tube diameter	d_1	40 mm
Loss coeff. for tube	β	12
First VT resonance	f_1	1.5 kHz
Second VT resonance	f_2	2.5 kHz
Q of VT resonances	Q	10
Charact. impedance	$(L/C)^{1/2}$	4 MPa s m^{-3}
Glottal open area		20 mm^2
Effective glottal length		5 mm
Blowing pressure in mouth	p_0	1 kPa
Max. area of lip opening	s_0	50 mm^2
Frequency rel. first resonance		1.03

As a first approximation, the small-signal result (27) can be used to predict the behavior of formants for such a cylindrical didjeridu. The result will not be accurate at low frequencies and applies to the high-tongue case in which the upper formants are emphasized at the expense of the drone fundamental. Because the vocal tract is a complex structure, this will be approximated by assuming that it possesses two resonant impedance minima above 1000 Hz, in this case at 1500 and 2500 Hz, respectively, these being separated by an impedance maximum. The tract can therefore be simply modeled as two series-resonant circuits in parallel, as shown in Fig. 1(d), with characteristic impedances appropriate to the cross section of the constricted upper part of the vocal tract. The third parallel inpedance represents that of the glottis and lungs. The playing frequency was taken to be half a semitone (3%) above the fundamental resonance of the tube, this being fairly characteristic of actual playing in the case of emphasized formants. The other relevant parameters are given in Table I.

The results of this calculation are shown in Fig. 3 and agree very well with high-tongue spectra measured on

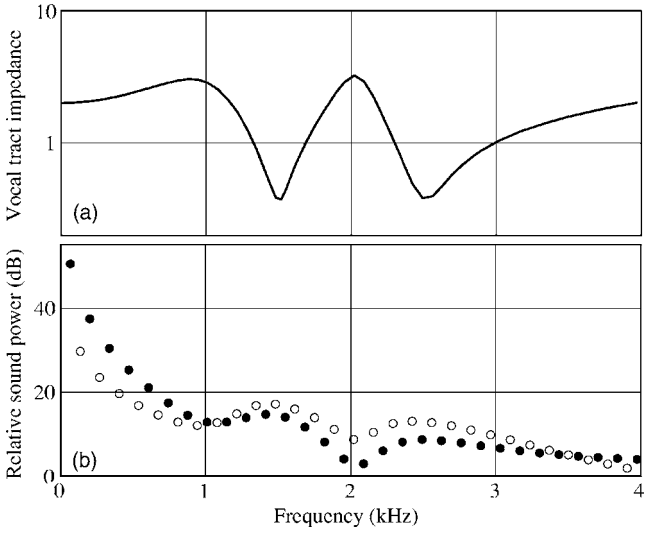


FIG. 3. (a) Vocal-tract impedance at the lips in units of MPa s m^{-3}, with two simple resonances at 1.5 and 2.5 kHz, respectively. (b) Computed radiated spectrum for the case of a cylindrical didjeridu tube with other parameters as in Table I. Odd harmonics of the drone frequency are shown with solid circles and even harmonics with open circles.

skilled players, as shown in Fig. 6 of our experimental paper,[13] which also used a cylindrical-pipe didjeridu. It is clear that there are two pronounced formants in the radiated sound spectrum, and that their center frequencies coincide with those of the two minima in the vocal-tract impedance. There is also a pronounced minimum at the frequency of the maximum in the tract impedance, as predicted. There are other interesting features of the spectrum illustrated in Fig. 3(b). Odd harmonics of the drone frequency are shown as filled circles and even harmonics as open circles. In the low-frequency range, odd harmonics are emphasized because of the maxima in the transfer function Z_{22}^D/Z_{12}^D close to those frequencies, but the fact that the drone frequency is a little above the first resonance means that the odd harmonics move progressively away from the transfer maxima, while the even harmonics move to closer matching and are consequently emphasized. An actual transfer of emphasis takes place at about 1000 Hz and another one in the reverse direction at about 3500 Hz. This effect is also clearly seen in experimental measurements on a cylindrical didjeridu.[13] In particular cases where there is very near coincidence between harmonic frequencies and tube resonances in a particular spectral region, this effect may even lead to the production of an emphasized formant band of completely different origin.

In the calculation shown in Fig. 3, the playing frequency was chosen to be 1.03 times the first tube resonance, which is typical for high-tongue playing. There is in this case no exact coincidence between an upper harmonic and a tube resonances and therefore no single enhanced harmonic. A very skilled player can, however, adjust the playing frequency to produce such a coincidence and therefore a greatly enhanced amplitude for a particular harmonic, much as is done in the "throat singing" of Mongolia,[37] though this technique is not used in traditional playing.

A rather similar simplified analysis can be applied to the case of a didjeridu with a flaring bore, though the algebra is necessarily a little more complicated. The results are generally similar to those demonstrated above. Because, however, the frequencies of the low-frequency tube modes do not have a simple harmonic relationship, there are no clear frequency regions in which even or odd harmonics dominates. In addition, the low harmonics above the fundamental are generally weaker because their frequencies do not approximate those of tube resonances.

VI. VOCALIZED SOUNDS

For some of the sounds traditionally made on the didjeridu, the player vibrates his vocal folds as in singing, as well as maintaining the lip vibration. This technique is traditionally used to represent the cries of animals in a performance that is illustrating a particular story or describing a place. It is important to examine the nature of the resultant sound.

In the simplest case, the performer plays the drone at frequency ω_D and sings a note of rather higher frequency ω_V. Each of the vibrating valves modulates the air flow at its operating frequency, so that it has a form like

$$U = A\langle \sin \omega_D t - x_1\rangle\langle \sin \omega_V t - x_2\rangle, \tag{28}$$

where x_1 and x_2 determine the closing fraction for each valve and the notation $\langle \ldots \rangle$ is meant to imply, as before, that only positive values of the quantity concerned are retained. The result is rather complicated and depends upon relative phasing of the two vibrations, but it is clear that the resulting flow will contain terms of essentially all frequencies $m\omega_D \pm n\omega_V$ where m and n are integers. The dominant terms will be those in which the values of m and n are small.

A particular example is the case in which the sung note is a musical fifth (frequency ratio 3:2) or more usually a musical tenth (frequency ratio 5:2) above the drone fundamental. The difference combination with $m=n=1$ in the first case or $m=2$, $n=1$ in the second then generates an impressive subharmonic drone of frequency $\omega_D/2$, together with all its harmonics.[13] Most of the other sounds used in performance are complex and transient, so that little is to be gained by analyzing them in detail.

VII. DIDJERIDU QUALITY

Although the geometrical form of native didjeridus varies widely, and good players can produce striking performances on a simple length of cylindrical PVC pipe, these players are able to rate didjeridus in terms of quality. There do not appear to have been extensive studies of these ratings, but beginning attempts have been made.[5–7] Amir,[6] in particular, has produced a quality criterion based upon the levels of harmonics 2 to 5 relative to that of the fundamental, a didjeridu that is rated excellent by players having low levels for these harmonics.

Examination of the didjeridu preferences of different clans in Arnhem Land shows, however, that there are striking differences in the preferred didjeridu shape.[38] Clans from Eastern Arnhem land, including the Yolngu, adopt a playing style in which the use of rhythmic articulation and higher tube modes for accents is pronounced, and their preference is for almost cylindrical didjeridus. Clans from Western Arnhem Land, in contrast, make little use of these higher modes but exploit the timbre changes possible through the use of vocal-tract resonances, and their preference is for didjeridus with a flaring bore profile. Nontraditional didjeridu performers use a combination of both styles, since they give didjeridu sound a quality that is not heard in any other musical instrument.

It is easy to see the reasons for these preferences. A narrow tube with a nearly cylindrical bore will have strong lower resonances in nearly harmonic relationship and it will be relatively easy to produce the higher modes. The drone, too, will be strong in quality because the well-aligned resonances will reinforce the first few odd harmonics, a point that is important since the sensitivity of human hearing is reduced at the low frequency, 60–70 Hz, of the drone fundamental. The strong quality of the drone fundamental and its lower harmonics then makes less obvious the subtle timbre changes produced by formant variations in the higher harmonics. These physical consequences appear to explain the preferences of those playing in the Eastern Arnhem Land style.

The reverse is true for the Western Arnhem Land style and preferences. In a widely flaring tube, overtones other than the first are rather difficult to produce because of the rapidly decreasing quality factor of the tube resonances, while the misalignment of these lower resonances means that they do not reinforce the lower harmonics of the drone, making it much less prominent than in the Eastern style. This then has the consequence that formant bands produced by vocal-tract resonances, which do not rely upon reinforcement by the resonances of the instrument bore, are prominent in the total sound as desired.

VIII. CONCLUSIONS

The primary purpose of the present paper was to examine the relation between maxima and minima in the vocal-tract impedance and the formant bands observed as a prominent characteristic of expert didjeridu performance. While the model still requires development if it is to give an accurate quantitative description of the radiated spectrum, both the theoretical development and the computed spectra demonstrate a close relation between the emphasized formant bands in the sound spectrum and impedance minima of the vocal tract, as measured at the lips. Direct evidence for this correlation is presented in the accompanying experiment paper.

ACKNOWLEDGMENT

This work was supported by a grant from the Australian Research Council Discovery Program.

[1]N. H. Fletcher, "Australian Aboriginal musical instruments: The didjeridu, the bull-roarer and the gumleaf," Proc. Stockholm Music Acoustics Conf. 2003, SMAC-03, pp. 201–204.
[2]T. A. Jones, "The didjeridu," Studies in Music (University of Western Australia) **1**, 23–55 (1967).
[3]T. A. Jones, "The yiraki (didjeridu) in Northeastern Arnhem Land: Techniques and styles," in *The Australian Aboriginal Heritage*, edited by R. M. Berndt and E. S. Phillips (Australian Society for Education through the Arts, in association with Ure Smith, Sydney, 1973), pp. 269–274.
[4]K. Neuenfeld, Ed., *The Didjeridu: From Arnhem Land to Internet* (John Libbey, Sydney, 1997).
[5]N. Amir, "A study of the didjeridu: Normal modes and playing frequencies," Proc. Internat. Symp. Musical. Acoust. ISMA-01, Perugia, Italy, 2001, pp. 95–98.
[6]N. Amir, "Harmonics: What do they do in the didjeridu?," Proc. Stockholm Music Acoust. Conf., SMAC-03, Stockholm, Sweden 2003, pp. 189–192.
[7]R. Caussé, G. Baudouin, and B. Sluchin, "An investigation on 'tonal' and 'playability' qualities of eight didgeridoos, perceived by players," Proc. Internat. Symp. Musical. Acoust. ISMA-04, Nara, Japan 2004, pp. 261–264.
[8]N. H. Fletcher, "Acoustics of the Australian didjeridu," Australian Aboriginal Studies **1**, 28–37 (1983).
[9]N. H. Fletcher, "The didjeridu (didgeridoo)," Acoust. Aust. **24**, 11–15 (1996).
[10]G. C. Wiggins, "The physics of the didgeridoo," Phys. Bull. **39**, 266–267 (1988).
[11]L. Hollenberg, "The didjeridu: Lip motion and low frequency harmonic generation," Aust. J. Phys. **53**, 835–850 (2000).
[12]N. Fletcher, L. Hollenberg, J. Smith, and J. Wolfe, "The didjeridu and the vocal tract," Proc. Internat. Symp. Musical. Acoust. ISMA-01, Perugia 2001, pp. 87–90.
[13]A. Z. Tarnopolsky, N. H. Fletcher, L. C. L. Hollenberg, B. D. Lange, J. Smith, and J. Wolfe, "Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I. Experiment," J. Acoust. Soc. Am. **119**, 1194–1204 (2006).
[14]A. Tarnopolsky, N. Fletcher, L. Hollenberg, B. Lange, J. Smith, and J. Wolfe, "The vocal tract and the sound of a didgeridoo," Nature (London) **436**, 39 (2005).
[15]L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1986), Chap. 3.
[16]N. H. Fletcher, *Acoustic Systems in Biology* (Oxford University Press, New York, 1992), Chaps. 8 and 10.
[17]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Secs. 8.2, 8.3, 8.15.
[18]E. R. Weibel, *Morphometry of the Human Lung* (Springer, Berlin, 1963).
[19]K. Ishizaka, M. Matsudaira, and T. Kaneko, "Input acoustic-impedance measurements of the subglottal system," J. Acoust. Soc. Am. **60**, 190–197 (1976).
[20]A. C. Jackson, B. Suki, M. Ucar, and R. Habib, "Branching airway network models for analyzing high-frequency lung input impedance," J. Appl. Physiol. **75**, 217–227 (1993).
[21]S. Mukai, "Laryngeal movement during wind instrument play," J. Otolaryngol. Japan **92**, 260–270 (1989).
[22]S. Mukai, "Laryngeal movement while playing wind instruments," Proc. Int. Sympos. Musical Acoust. ISMA-92, Tokyo, Acoust. Soc. Japan 1992, pp. 239–242.
[23]J. Wolfe, J. Smith, G. Brielbeck, and F. Stocker, "A system for real time measurement of acoustic transfer functions," Acoust. Aust. **23**, 19–20 (1995).
[24]J. Epps, J. R. Smith, and J. Wolfe, "A novel instrument to measure acoustic resonances of the vocal tract during phonation," Meas. Sci. Technol. **8**, 1112–1121 (1997).
[25]J. Wolfe, A. Z. Tarnopolsky, N. H. Fletcher, L. C. L. Hollenberg, and J. Smith, "Some effects of the player's vocal tract and tongue on wind instrument sound," Proc. Stockholm Music Acoust. Conf., SMAC-03, Stockholm, Sweden, 2003, pp. 307–310.
[26]N. H. Fletcher, *Acoustic Systems in Biology* (Oxford University Press, New York, 1992), Appendix B.
[27]H. F. Olson, *Acoustical Engineering* (Van Nostrand, New York, 1957), pp. 100–123.
[28]S. Yoshikawa, "Acoustical behavior of brass player's lips," J. Acoust. Soc. Am. **97**, 1929–1939 (1995).
[29]S. Adachi and M. Sato, "Trumpet sound simulation using a two-dimensional lip vibration model," J. Acoust. Soc. Am. **99**, 1200–1209 (1996).
[30]D. C. Copley and W. J. Strong, "A stroboscopic study of lip vibrations in a trombone," J. Acoust. Soc. Am. **99**, 1219–1226 (1996).
[31]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Sec. 13.3.
[32]N. H. Fletcher, "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180 (1993).
[33]A. Tarnopolsky, N. H. Fletcher, and J. C. S. Lai, "Oscillating reed valves—An experimental study," J. Acoust. Soc. Am. **108**, 400–406 (2000).
[34]N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Gordon and Breach, New York, 1961).
[35]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Chap. 5.
[36]J. Gilbert, J. Kergomard, and E. Ngoya, "Calculation of the steady-state oscillations of a clarinet using the harmonic balance technique," J. Acoust. Soc. Am. **86**, 35–41 (1989).
[37]S. Adachi, "An acoustical study of sound production in biphonic singing, Xöömij," J. Acoust. Soc. Am. **105**, 2920–2932 (1999).
[38]S. Knopoff, "Accompanying the dreaming: Determinants of didjeridu style in traditional and popular Yolngu song," in *The Didjeridu: From Arnhem Land to Internet*, edited by K. Neuenfeld (John Libbey, Sydney, 1997), pp. 39–67.

Lateral distribution of ultrasound velocity in horizontal layers of human teeth

Christoph John[a)]
Biomedical Engineering Laboratory, Department of Conservative Dentistry, University of Tuebingen, Osianderstr. 2-8, D-72076 Tuebingen, Germany

(Received 14 April 2005; revised 30 September 2005; accepted 4 October 2005)

The speed of ultrasound at 20 MHz differs inside human teeth depending on which tissues are involved. At least two out of four dental tissues exhibit variations in the longitudinal velocity (c_L). The aim of this *in vitro* study is to describe the laterally varying propagation velocity of tangentially propagating longitudinal waves. At a distance of 5 mm from the crown reference, the c_L is determined using longitudinal sections and a pulse-echo technique. Several graphs are combined to account for the corono-apical decrease in c_L and the laterally varying c_L distribution along horizontally adjacent relative tooth width portions. The laterally increasing c_L of 21 specimens at radial locations rises from 2900 to 4000 m/s. A mathematical evaluation reveals an optimal horizontal formula of the form $c_{L(5\ \mathrm{mm})}=a+bX^2\ln(X)$, where X is the standardized lateral parameter relative to individual tooth width w, which is compensated for offsets. Individual residuals and a,b coefficients of the corresponding approximations are provided. Individual mean errors range from 7 m/s (SD=6 m/s) to 92 m/s (SD=79 m/s). The lower contour of the envelope curve of all c_L distributions is described by taking up a formerly introduced equation [J. Acoust. Soc. Am. **116**, 545 (2004)]. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2146083]

PACS number(s): 43.80.Ev, 43.35.Cg, 43.80.Jz, 43.20.Jr [FD] Pages: 1214–1226

I. INTRODUCTION

Human teeth comprise a living soft tissue and three distinct mineralized tissues: the enamel, the cementum, and the dentin. The anisotropic enamel layer and the thin cementum layer are covering the crown and the root portions of the tooth, respectively. The dentin core beneath them shelters the endodontic system and constitutes the largest part of the human tooth.[1] The individual thickness of the dentin changes depending on the type of tooth, the directions measured, and the age of the patient.[2–4] When comparing some properties of sound dentin and human bone, like density,[5,6] the mechanical properties,[7,8] and the ultrasonic velocities,[6,9,10] they seem remarkably similar. A local similarity can be explained by compatibility considerations at the apical region of the root.[11] Since bone structures model or remodel themselves according to the functional requirements imposed on them, and because teeth act as a mechanical device during masticatory processes, it is considered whether any adaptation is observed for histologically more complex teeth that are dimensionally much smaller and exhibit a corono-apically tapered shape. The dentin alters its porous structure, its permeability, and its local degree of mineralization as a function of age and depth.[12–16] Its internal structure is most obviously influenced by the convergence and the great amount of dentinal tubules that spatially vary in size and concentration.[1,14] The radial geometry of tubules[17] reveals some deviation from the main tubule orientation because the conical tubules are not straight and also because some tubule lumens are interconnected.[1] Mainly, the radial enlargement of the conical tubules towards the pulp play a key role for the speed of adaptation and the degree of mineralization. Apical dentin of the root seems to be more frequently sclerosed and reveals a higher mineralization.[15] Featuring these aspects and the underlying structural arrangements one may most likely assume a spatial variation in its acoustical properties.[9,18,19] Data about the frequency distribution of density are available[20] and variations in the ultrasonic velocity of hard dental tissues as well as the acoustic impedances are already investigated.[9,21] In general, any variations in the ultrasound parameters complicate the diagnostic usage of ultrasound in dentistry, especially if these variations are unknown or high. Therefore, a complete quantification of visco-elastic parameter distributions is helpful or may even be considered as a prerequisite for utilizing ultrasonographic probes as diagnostic tools that localize any invisible internal damages, too.[22] Recently, a partial quantification of the vertical distribution of sound velocity approximates the initial corono-apical decrease in the longitudinal wave velocities (c_L) of 21 dentin specimens of extracted human teeth by $c_L=4224-(257\ln(y))$.[9] The former data are based on a single relative tooth width portion and a standardization that accounts for a distinct number of roots ($x=0.7w$ for teeth with a single root and a substitute of $x=0.826w$ for teeth with two or more roots). In the present paper, multiple relative tooth width portions $x=(m0.02)w$ are evaluated, corono-apical height levels h_y are accounted for, and adjacent points are measured at several distinct locations relative to the tooth width (i.e., as a function of relative radial distance or relative width X). This is performed to simultaneously describe the corono-apical decrease in cervical c_L and the laterally varying c_L along horizontal x coordinates.

a)Electronic mail: cjohn@ieee.org

II. MATERIALS AND METHODS

A. Specimens

Twenty-three specimens are used for measuring the speed of ultrasound at 20 MHz. Vital human teeth are utilized that are expected to be caries and restoration free. All teeth are extracted for orthodontic or periodontic reasons. After mechanical cleaning, the teeth are stored in jars of polypropylene that are filled with a physiological saline solution. The 15 ml jars (Straight-Side Wide-Mouth Jar, Nalgene™ Labware® Brand Products, Sybron Corp./Nalge Company, Rochester, NY, USA) exhibit a measured diffusion rate that contributes to a continuous loss of water of about 0.5 g/m^2 per week. This rate is observed at an average room temperature of 23 °C and for an average external humidity of 34% versus the internal high humidity.

All vital human teeth are embedded in a commonly available resin (Technovit©; Kulzer, Hamburg, Germany). One ground section from each of 21 teeth is obtained from their central sagittal areas ("$x_i y$" and "$x_j y$" planes in Fig. 1) and two horizontal cross sections containing the L_5-plane (Fig. 1) are taken from the remaining two teeth. The thicknesses of the horizontal sections are 1.5 and 1.8 mm, respectively. Each longitudinal section of thickness $d = 1.0 \pm 0.44$ mm (Fig. 1) is prepared using a low-speed sectioning machine, equipped with a thin diamond disc of thickness 0.15 mm (Isomet© and 15HC 11-4243; Buehler, Lake Bluff, IL).[4,9] Utilizing a mounted glass fixative, it is possible to control and continuously readjust the blade orientation during the cutting process for the purpose of achieving coplanar sections. In addition, care is taken to adequately select the best sectioning parameters, which significantly reduce any ripping pattern of the section surfaces. The individually greatest deviation in section thickness varies between 10 and 254 μm. This corresponds to deviations from a perfectly coplanar section by angles that usually are not greater than 0° 19′ (the outsized sample cut by an old wavy blade shows 0° 48′).

B. Descriptive data

1. Anamnestic data

The ages of the patients from which the teeth are extracted after a specific time period of oral function are an important aspect one may have to account for when comparing the distributions of all c_L data. One quarter of the specimens came from patients that are younger than 25 years, one quarter is older than 47 years, and the age where the group is split into two subgroups at the median is 40 years.

2. Maximal tooth widths $w_{il/j}$

The maximal width of all teeth varies from 5.0 to 10.3 mm, with a mean of 7.3 ± 1.7 mm. All but one tooth exhibit the maximal width at their crowns. More detailed data about the distribution of the widths are given in Table I. The variation in "crown width" is investigated as a function of the type of tooth.

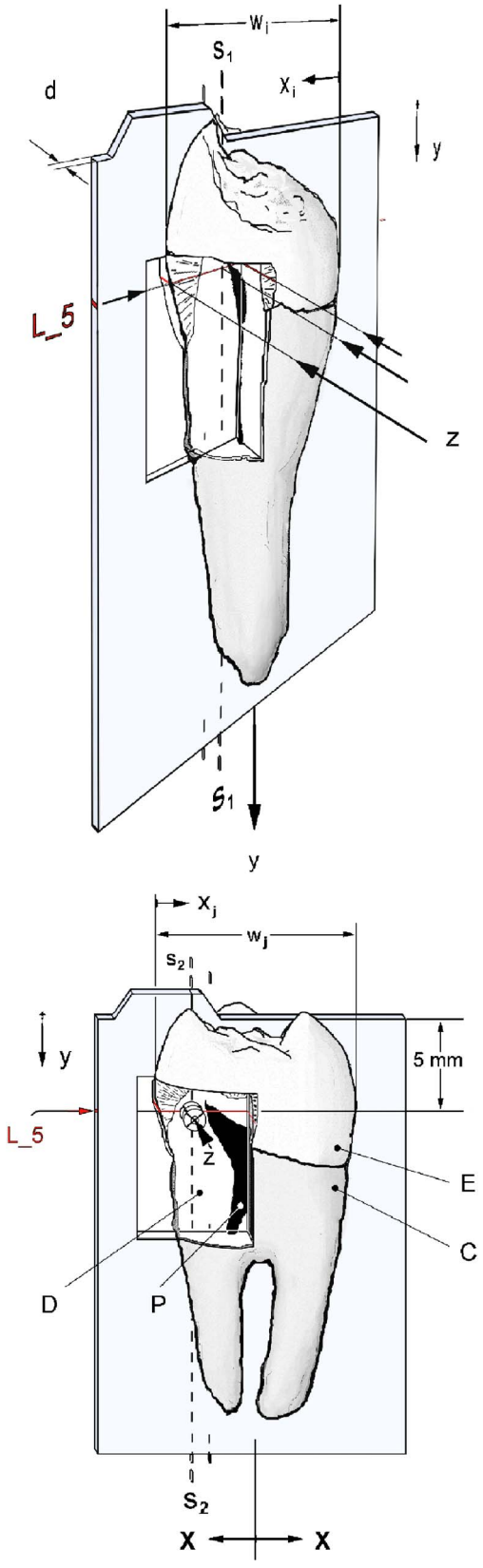


FIG. 1. (Color online) Unified approach for the measurement of *p*-wave velocities along adjacent "*z*" directions in distinct types of human teeth. Horizontal scans at 5 mm from the crown reference ("*L*_5") are used to determine the lateral distribution of ultrasound velocity outside the pulp chamber boundary ("*P*"). Two longitudinal sections of thickness "*d*" are distinguished (a) with a single root (top) and (b) with two roots (below). The dentin core ("*D*") is physiologically covered by a thin cementum layer ("*C*") and an enamel cap ("*E*") at the roots and the crown, respectively.

TABLE I. The crown width of human teeth varies depending on the type of tooth and the number of roots. Molars are the widest teeth.

Type of tooth	Crown width (mm)
Incisors (I)	6.0±1.2
Premolars$_{\text{Single Root}}$ (P_1)	6.3±0.2
Canines (C)	7.0±1.1
Premolars (P)	8.0±1.6
Premolars$_{\text{Two Roots}}$ (P_2)	9.2±0.2
Molars (M)	10.0±0.3
Total	7.3±1.7

3. Relative width

The maximal width is abbreviated by w_i when representing teeth with a single root (Fig. 1, top) and by w_j when representing teeth with two roots (Fig. 1, bottom). Each individual absolute w parameter is numerically divided into equi-proportionate relative tooth width portions. The center of the tooth is assigned to the relative tooth width position $x_c=0.5w$. A radially oriented parameter X_r and an offset compensation X_c are introduced. X_r is defined by the absolute of the difference between the actual relative x of an arbitrary position $p(x,y)$ and the individual center coordinate x_c according to the following term:

$$X_r = |x/w - 0.5|. \tag{1}$$

This way, any lateral position close to the pulp may be uniformly represented by expressions similar to $x=0.46w$ and a radially oriented parameter $X_r=0.04$ (or $x=0.6w$ and $X_r=0.1$). Those positions at the outer surface of the tooth are represented by values close to $x=w$ or $x=0$, and the uniformly standardized relative width $X_r=0.5$.

Modified values result when a normalizing procedure is applied that renders any original value into a standardized "relative width" according to

$$X = |x/w - 0.5| - X_c \tag{2}$$

with $X_c=0.126(N-1)$ and where X_c partially compensates for the extremes of the pulp chamber geometries found in sections with a distinct number of visible roots N.

C. Experiments

Preliminary experiments are conducted for two horizontal cross sections that are irradiated parallel to y (Figs. 1 and 2). The main experiments are conducted for 21 coplanar ground sections. Using ultrasonic energies at normal incidence, these sections are also perpendicularly irradiated. The corresponding axes of the ultrasonic beams are aligned with the z directions illustrated in Figs. 1–3. (The lower image portion of Fig. 1 depicts z as being almost orthogonal to the paper plane.) Depending on how many roots a specimen has, the varying c_L is determined by means of distinct coronoapical line scans (Fig. 1: S_1 at distance x_i and S_2 at distance x_j, respectively). Each scan starts at the enamel crown reference point and runs through the dentin core. They pass the height level at 5 mm from the crown reference and end at the apical cementum layer. The X position of the line scans is varied. Finally, horizontal scans depict the laterally varying c_L at that specific height level, where the marker denotes L_5 (Fig. 1).

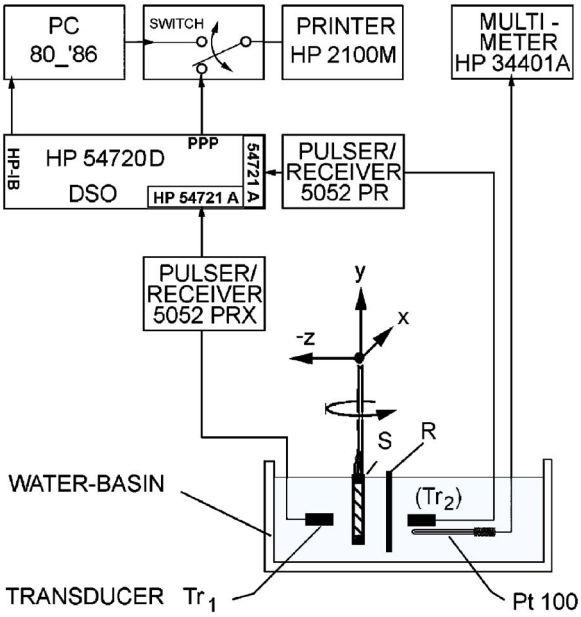


FIG. 2. (Color online) Sketch of the experimental setup showing the components of the ultrasonic measuring system. Radio frequency pressure pulses are generated and received echograms are amplified by means of a pair of spherically focused 20-MHz transducers and a pair of pulser/receivers. The orientation of the coordinate system is shown with respect to the tooth section ("*S*") and the rotating reflector ("*R*").

In order to summarize the lateral variation of the p-wave velocity in tangentially irradiated dentin, the c_L at 20 MHz is recorded and illustrated by

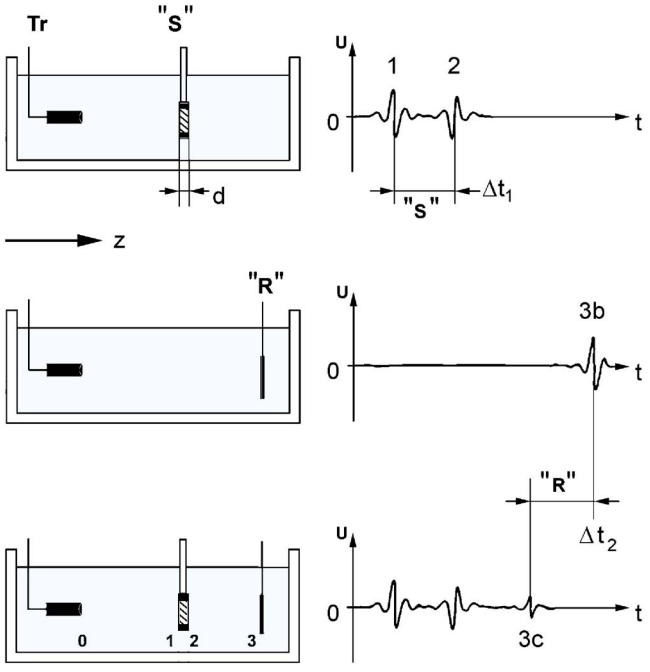


FIG. 3. (Color online) The determination of acoustic p-wave velocities in longitudinal sections ("*S*") of human teeth is based on a measuring principle that compares relevant differences in times-of-flight ("Δt_1" and "Δt_2").

(a) many adjacent, equidistantly measured, vertical locations in every single specimen (Appendix A) and
(b) the spatial distribution within a single height level in the L_5 plane (Appendix B).

For the purpose of providing adequate descriptions, typical locations are standardized as described in Sec. II B 3. Furthermore, several procedures for approximation are utilized, the quality of the fits are determined, residuals are computed and economical aspects are considered in Appendixes A and B.

During the fitting of the measured c_L data, the set of possible functions is condensed with respect to finding the most valuable ones. In general, a new "function of priority" is considered to be more appropriate than others provided that it is simple and accurate (Appendix C). In order to integrate data in an efficient manner into models designed for medical element engineering and simulation (MEES) a total of 11 "simple-and-accurate" criteria are evaluated by accounting for different weighing factors attributed to each criteria (Appendix D).

D. Measurement arrangement

Ultrasonic pressure pulses are generated by the left 20-MHz transducer (Fig. 2). Both immersion-type transducers are spherically focused (V317; Panametrics, Waltham, MA) and are attached to the high-voltage outputs of two pulser/receivers (5052 PRX; Panametrics, Waltham, MA). The whole measurement arrangement comprises the transducers (Tr_1,Tr_2), the tooth specimen (S), several electronic devices, a Platinum 100-sensor, a water basin, and the removable copper reflector (R) (Fig. 2). The xyz table allows for movements at reproducible steps. The Pt-100 sensor and the GPIB-bus-multimeter (HP 34401 A; Hewlett Packard, Colorado Springs, CO) are used to measure the equilibrium temperature excursion during the immersion experiment. Further details about the components of the equipment, the lateral resolution, the axial resolution, and the modifications of former arrangements can be found elsewhere.[9,18,23–26]

E. Measurement principle

The velocity is measured using a modified pulse-echo method. This is performed according to three steps (Fig. 3).

When an ultrasonic pressure pulse propagates in a medium, the distance traveled by the pressure pulse, the ultrasonic velocity, and the time-of-flight are coupled together. Provided that two of them are known, the third can be calculated. However, in the present study as well as in several preceding investigations it is advantageous to determine the speed of propagation without knowing the distance through which the sequence of pulses propagates.[9,18,23–25,27–29] This is accomplished by applying a combination of the reflection method (Fig. 3, top) and the substitution method (Fig. 3, middle and bottom). The p-wave velocity in human dentin is then derived from the quotient of Δt_2 (substitution method) and Δt_1 (reflection method) which is used as an increase factor relative to the measured longitudinal velocity in water.[30] Except as explicitly stated for Fig. 4, and parts of Fig. 5 where $\mathcal{N}=21+2$, all illustrated ultrasonic velocity data are based on computations that account for 21 longitudinally sectioned specimens.

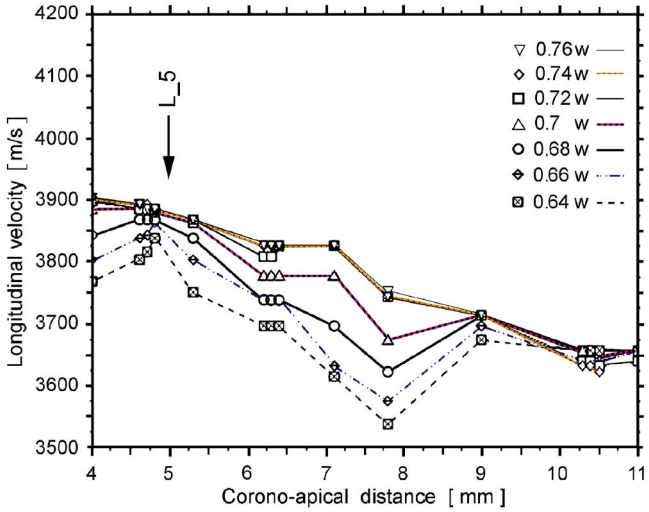


FIG. 4. (Color online) Tooth-width-dependent diagram showing longitudinal p-wave velocities that vary corono-apically along the long axis of the tooth for seven adjacently arranged line scans in one individual specimen. Please note that a laterally varying x_{jli} range of [0,…,1.0w] corresponds to a radially varying "relative width" X of [0,…,0.5]. The position of the L_5 plane is indicated.

III. RESULTS

A. Lateral variation of the corono-apical decrease in ultrasound velocity

1. Teeth with several roots

The results of all specimens that belong to teeth with two or more roots are superimposed to the majority of teeth with a single root. This is accomplished by subtracting a value of $X_c=0.126$ from the corresponding width portions according to Eq. (2). This correction may be interpreted as a linear shift outwards or away from the tooth axis.

2. All teeth

The corono-apical decrease in the p-wave velocity is illustrated for a window that covers a certain vertical tooth portion (Fig. 4). This tooth portion lies in between 4 and 11 mm from the crown reference point (Fig. 4). For each

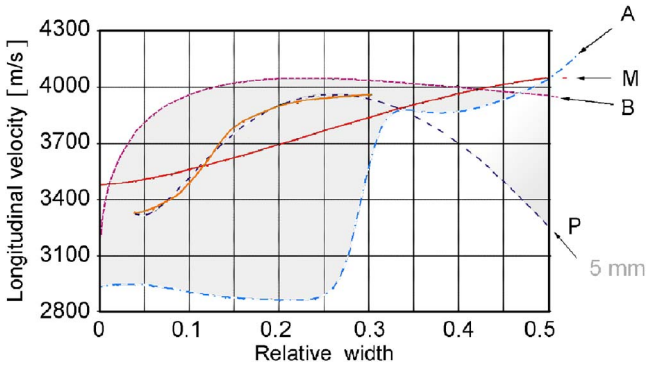


FIG. 5. (Color online) The lateral distribution of the observed c_L at the level of L_5.

TABLE II. A multi-purpose term (see Ref. 9) of the form $(\Sigma(a_i y^i))/(1+\Sigma(b_{i+1} y^{(i+1)}))$ can be used to relate the Young's modulus (E), the density (ρ), and the p-wave velocity (c_L) with the Poisson's ratio (ν_0) according to Eq. (4) and (6). The laterally varying c_L at 5 mm is described by a distribution pattern where all values are above a lower envelope function $A_{L\_5}(X) = c_{L\ (L\_5)}\text{min}$.

	q (Eq. (4)) $=E/(c_L^2\rho)$ $=f(\nu_0)$	$c_{L\ (L\_5)}\text{min}$ $=A_{L\_5}(X)$
Valid X, ν_0	0.0,…,0.5	0.0,…,0.5
Fit.Std.Err.	0.0006	0.003
R^2	>0.999	>0.999
a_0	1.0	2933.6
a_1	0	−36 246.33
a_2	−4.00	265 980.65
a_3	0	−1 008 554.2
a_4	0	1 410 240.3
	1	1
b_1	0	−12.654
b_2	−1.662	98.475
b_3	0	−407.7098
b_4	1.5406	702.1173
b_5	0	−284.834 72

scan line S_h that is adjacent to the next parallel scan S_{h+1} the assorted length where a continual decrease in c_L can be observed is increasing with increasing normalizing relative tooth width locations (from inside to outside, i.e., from $0.64w$ to $0.76w$). Each scan reveals a decrease in c_L from coronal portions (i.e., near 4 mm) to apical portions (i.e., near 11 mm). The common distance where the continual decrease can be observed for all normalizing relative tooth width portions starts uniquely at a crown distance of about 4.75 mm and ends near 7.75 mm (Fig. 4). An inserted arrow is indicating the position of the horizontal L_5 data plane, which is introduced to explicitly show and compute quantitative results of an interesting quasi-horizontal layer. Please note that the largest decay along the 3-mm distance is about 300 m/s. This corono-apical decrease in c_L is found rather centrally (e.g., at $0.64w$, i.e., between 3830 and 3530 m/s). Smaller decays are found in a consecutive manner and almost continuously when going to investigate more peripheral regions (e.g., 225 m/s at $0.76w$ in Fig. 4).

B. Lateral distribution of tangential c_L (L_5 plane)

When plotting all c_L data of the L_5 plane as a function of their original location, the relative width is utilized as a unified and standardized locator, which determines where to uniquely compare tooth specimens with distinct maximal crown widths and a distinct number of visible roots. Figure 5 summarizes the c_{LL_5} distribution accordingly. The lateral distribution of the observed p-wave velocities at height levels of 5 mm are shown.

All data are summarized by (a) the minimum and the mean ("M") of all specimens (Table II and III B 4), (b) a typical "5 mm" curve that looks as if it belongs to a data cluster inside a rotated hysteresis curve, (c) a total of 14 individual approximations of all laterally varying velocities (Table III and Sec. III B 5), (d) their residual errors (Table IV), (e) the cluster's lower envelope function ("A" and Sec. III B 1), and (f) the cluster's upper envelope function ("B" and III B 2). Numerical data for A are provided by adequate coefficients of an already introduced multi-purpose equation (C_{L_5}min in Table II, right column).

TABLE III. Lateral distribution of c_L. The lateral behavior of the longitudinal ultrasound velocity in human dentin at a vertical level of 5 mm ($c_{L\,L\_5}$) relative to the crown reference is approximated by a function that utilizes the lateral variable $X=|x/w-0.5|-X_c$.

	Approximation at level $y'=5$ mm $c_{L\,L\_5}=a_5+b_5\ X^2 \ln(X)$		
$X=(\lvert x/w-0.5\rvert-X_c)$ Specimen	a_5 (m/s)	b_5 (m/s)	R_5^2
M	3706	−1954.3	0.97
E	3251	−6711.4	0.96
R	3072	−6410.6	0.93
J	3088	−7286.7	0.93
N	2748	−8698.9	0.92
F	2363	−17151	0.91
O	2759	−8659.6	0.91
I	3083	−10397	0.90
C	3380	−6592.5	0.86
A	3518	−2296.1	0.86
G	3633	−3359.0	0.85
S	3206	−6927.1	0.84
3 ($a\_\max, b\_\min$)	3778±33	−1053±246	0.74±0.09
6 Remains	3282±388	−7648±3572[a]	0.63±0.12[b]

[a]Exhibiting no unified gradual behavior.
[b]Reflecting local exceptions from relevant smooth continuity.

1. Minimal envelope curve

The minimal envelope curve is computed and is illustrated in Fig. 5 by the lower contour graph (A). The best A-fitting is based on Padé shape functions. The approximation utilizes rational polynomials (RPs), which are known from previous publications to allow for convex shapes and interpolating functions in real space.[9,31–33] The set of mathematical formulas that are computed to fit best the lowest c_L data according to Appendix B, Procedure III (a) can be preferably described by a generated RP

$$c_{L(L\_5)}\ \text{min}(20\ ^\circ\text{C})/(\text{m/s}) = \frac{\Sigma_{i=0}^{4} a_i \cdot X^i}{\left(1+\Sigma_{j=1}^{5} b_j \cdot X^j\right)} \tag{3}$$

with $j=i+1$ and i ranging from 0 to 4, where X is the unitless variable of the standardized relative width. In homogeneous and isotropic media, the exact relationship between the longitudinal p-wave velocity (c_L) and the elastic behavior may be expressed in terms of the density (ρ), the Young's modulus (E), and the Poisson's ratio (ν_0) according to

$$(c_{L(L\_5)})^2 \cdot \rho = \frac{E}{(1+\upsilon_0)\cdot(1-2\cdot\upsilon_0)} \cdot \frac{1}{(1/(1-\upsilon_0))}. \tag{4}$$

With possible Poisson's ratios $0 \leqslant \nu_0 \leqslant 0.5$ one can find that the convergence criterion for

TABLE IV. Differences between $c^{*}_{L\,L\,5}$ measurements in dentin and the corresponding $c_{L\,L\,5}$ approximations at 5 mm.

	Individual errors at level $y'=5$ mm			
Specimen	(Residuals <0)$>$ (%)	(Residuals >0)$<$ (%)		Mean $\lvert$Residual$\rvert\pm$SD (m/s)
M	−0.5	+0.5		7±6
E	−1.9	+1.9		31±23
R	−2.1	+2.3		68±38
J	−2.3	+2.0		50±25
N	−2.3	+3.2[a]		72±48
F	−3.5[b]	+2.9[d]		90±62
O	−3.6[g]	+3.5		70±54
I	−0.5	+3.1		91±51
C	−2.9	+3.4		79±39
A	−1.3	+0.9		23±11
G	−1.5	+2.0		32±23
S	−3.0	+3.8[b]		80±43
K,*L*,*T*	−0.8	+1.0		13±10
6 Remains	−3.0[k]	+2.8[c]	}	92±79
	−3.8[j]	+3.9		
	−3.8[f,i]	+3.8[e]	}	
	−2.6[h]	+3.1[d]		
	−3.8	+2.2	}	
	−0.9	+0.7		

[a]Outlier (no perfect continuity, due to a single data point with 4.2% error).
[b]dto. (4.4%;
[c]4.5%;
[d]4.8%;
[e]4.9%;
[f]5.7%;
[g]6.0%;
[h]7.1%;
[i]7.2%;
[j]7.8%;
[k]12.9%).

$$\lim_{F\to\infty}\sum_{h=0}^{F}\upsilon_0^h=\frac{1}{(1-\upsilon_0)} \tag{5}$$

is fulfilled for all Poisson's ratios because $|\nu_0|<1$. This means that the denominator of the last term on the right side of Eq. (4) can be reconsidered to generally contribute to $\Sigma_{h=0}^{F}\upsilon_0^h$. The corresponding first approximation error ($\mathrm{Err}_{\_F}$) depends on F and ν_0. It is found that even with $F=4$ (or more) this error $\mathrm{Err}_{\_F4}$ is not greater than 3.125% for all possible Poisson's ratios ranging from 0 to 0.5. For better convenience, both sides of Eq. (4) may be divided by E provided that there is some elasticity. Then, the inverse can be utilized to relate E, ρ, and c_L to the Poisson's ratio by adequately integrating the best replacement of $(1+\nu_0)$ and $(1-2\nu_0)$ provided that the thickness is very much smaller than the width of the specimen. Although the derivation of suitable equations and a potential review of linear viscoelasticity models of porous brittle materials with a microstructure is not the topic of this article, it should be mentioned that the product's equivalent of $(1-\nu_0-2\nu_0^2)=\Sigma_{i=0}^{2}c_i\upsilon_0^i$ can be combined with $\Sigma_{h=0}^{F}\upsilon_0^h$, or it can favorably be replaced by an adequate approximate term. Finally, the whole expression $(1-\nu_0-2\nu_0^2)\Sigma_{h=0}^{F}\upsilon_0^h$ can be rewritten by $(\Sigma_{h=0}^{F}\upsilon_0^h)\Sigma_{i=0}^{2}(c_i\upsilon_0^i)$ and by similar polynomials. Multiplying the polynomial by unity via a numerator term and the same nonzero denominator term, any polynomial can be transformed into any RP. The advantage of the first step ($\mathrm{Err}_{\_F}$) is the ability to introduce a predefined error contribution with respect to Poisson's ratios. The advantage of the last step (RPs) is twofold: The simultaneous opportunity arises to partially counterbalance only unwanted portions of the previously introduced first errors and to maintain a general up-grade compatibility. Considering expressions for quasi-elastic constants that depend on frequency, it is possible to compensate some error contributions depending on the viscoelastic dispersion law. The rheology is defined by the coefficients introduced. The compatibility featured by this type of equation may be helpful as long as it is not clear whether healthy dentin in the wet state can be modeled as an isotropic continuum or as a hexagonally anisotropic structure.[34–37] This way, distributions of the Young's modulus, the density, and the longitudinal *p*-wave velocity can be related to distributions of the Poisson's ratio by a sophisticated approximation that follows:

$$\frac{E}{(c_{L(L\_5)})^2\cdot\rho}=\frac{\Sigma_{i=0}^{4}a_i\cdot\upsilon_0^i}{1+\Sigma_{j=1}^{5}b_j\upsilon_0^j}+\mathrm{Err} \tag{6}$$

and where only four specific coefficients (a_0, a_2, b_2, and b_4) need to be used if relative errors (Err) of no more than 2.9 tenths of a percent can be accepted for observable Poisson's ratios ranging from 0 to 0.4. The similarity of Eq. (3) and Eq. (6) is intended. The computed coefficients of the approximate equation (6) are listed in Table II (left column). The computed coefficients of the best minimum fitting equation (3) are listed in Table II (right column). Both equations may serve for numerical evaluations of elasticity.

2. Maximal envelope curve

In order to further quantify the lateral distribution of the c_L data in the L_5 plane by additional means, the maximal envelope curve is computed and is illustrated in Fig. 5 by the upper contour graph (B). The best fit for B is given by a mathematical formula according to

$$B(X,20\ ^{\circ}\mathrm{C})/(\mathrm{m/s})=-3200+4760\cdot X^{0.5}+6250e^{-X}. \tag{7}$$

Additional details are referred to in Appendix B, Procedure III (b).

3. Tubule density and tubule orientation at intercept P

Most data points are found in the left two-thirds of Fig. 5 and especially rather close to the sinusoidal portion of the partially inserted Padé graph.[31–33] When this graph portion is approximated by another best approximation that is not limited by a continual growth of the graph, this leads to an intersection at P. Particular portions of the right third of the figure between P and B (gradually shaded area) are important for clarifying the interaction of the propagating compressional waves with the tissue's dentinal tubules and their

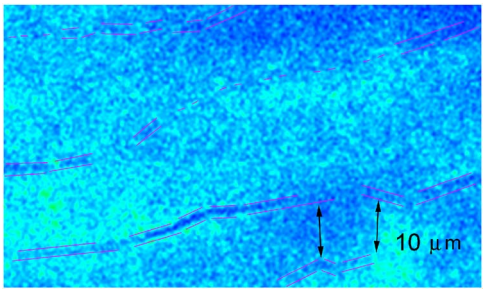


FIG. 6. (Color online) Magnified view of a peripheral portion of human dentin using confocal laser scanning microscopy (CLSM). The specimen is acoustically irradiated in a normal direction to the paper plane. A series of several depth sensitive images helps to identify the individual courses of the dentinal tubules (markers).

main orientation in cross sections. The two horizontal cross sections contribute to the data set in a rather distinct manner because the ultrasonic beam axes at P (and above P) are perpendicular to the specimen's surfaces and almost perpendicular to the orientation of the dentinal tubules. The tubule densities are generally the lowest for relative widths near $X=0.5$ (peripheral dentin). In the peripheral dentin a p-wave velocity of 3800 m/s is determined for a location where the regional amount of tubules covers a relative area of 6.6% (Fig. 6). At similar locations where the perpendicularly irradiated dentinal tubules exhibit a variation in tubule diameter of 1.0 ± 0.2 μm, the resulting acoustic interaction mechanisms do not inevitably cause any most extreme p-wave velocities in human dentin.

4. Mean horizontal distribution of c_L (L_5 means)

The laterally varying mean of individual c_L distributions is illustrated by (M) in Fig. 5. The approximation is based on procedure II (Appendix B). The formula with rounded values is

$$M(X,20\ ^\circ\mathrm{C})/(\mathrm{m/s}) = \frac{1}{\frac{1}{3508} - (X/3785)\sqrt{X} + (X^2/2841)\sqrt{X}}. \tag{8}$$

5. Spatial distribution of c_L (Individual c_L at L_5)

The individual spatial variations in c_L are based on procedure V (Appendix B) and the following formula:

$$c_{L_{(\mathcal{N}=1,\ldots,6)}}(X,20\ ^\circ\mathrm{C})/(\mathrm{m/s}) = a_5 + b_5 X^2(\ln X). \tag{9}$$

Equation (9) fits well along differently reduced distances in dentin (with X ranging from uniquely 0.02 to individual values between 0.4 and 0.5). Application of procedure I contributes to a variety of adjusted parameter pairs, which are listed in Table III. The variation of the individually fitted spatial distributions of c_L in dentin reveals a great inhomogeneity.

C. Errors

The error in measured velocity is less than a tenth of a percent and mainly depends on the general possibility to control and maintain a fixed temperature inside the water bath where the ultrasonic waves are propagating. The errors of the approximations performed according to procedure I correspond to the sum of the smallest quantity of a 32-bit real variable (negligible) and the computed residuals. The computed $\mathcal{N}\leqslant 6$ mini-group's residual errors and the individual residuals at any data points of all single approximations are all but one less than 175 m/s (one single extreme of 365 m/s). Residual means are found that vary according to the list provided in Table IV.

The overall-summary given by Eq. (8) is based on measured values at standardized locations, which are assumed to be precise with a spatial error of ±0.05 mm. However, an additional lateral shifting error is imminent because rounded values of X are involved that are integer multiples, like $m\times 0.02w$. The individual residual error in each data point can be as high as 13% (Procedure II, Appendix B) and 12.9% (Procedure V, Appendix B), with deviating values that are ranging from a minimum between −0.5% and −3.8% to a maximum between +0.5% and +3.9% (Table IV). The means of the squared roots of the residuals are listed and denote a given span ranging from 7 ± 6 m/s to a maximum of 92 ± 79 m/s.

IV. DISCUSSION

A. Molars

Since the place required for any additional root predominantly determines the dimensions of the tooth—including its width parameter—any categorizing split parameter (like the number of roots) has to be consistently applied for improving the statistical power of the analysis. Therefore, the variation in crown width is investigated as a function of the type of tooth and for one type of tooth as a function of the number of roots, too. This is done for premolars only, because the number of molars exhibiting a distinct number of roots is not large enough to make a categorizing split. Another problem one has to account for is how to get a longitudinal section that shows and covers three roots at the same time when the corresponding section thickness should be not too large but nevertheless similar to the rest of the specimens. This matters due to an incompatible compromise between the anatomy of the teeth and experimental requirements. Anatomy exhibits complex convex and concave curvatures at tooth surfaces. Experiments require quite coplanar sections with a measurable, rather constant thickness. The initial number of extracted molars ($n=60$) does not allow for a consistent split into groups of longitudinal sections with one, two, or more roots. During the course of the study the question arises whether sections with one or two visible roots can be properly assigned if they actually belong to any molars with three roots. Since this is principally impossible, any arbitrary (or best guess) assignment does not make sense. What is of equal importance is that such an improper assignment would also cause multiple echoes from nonplanar rough surfaces of sections whose cutting planes more hazardously (or inevitably) incorporate the complicating geometries of the endodontic root canal system of any third root.

When considering the L_5-plane one may want to distinguish between teeth with a single root and those teeth with two (or more) roots. Doing so, specimens with a single root

TABLE V. The modulus of elasticity is calculated for wave numbers from transformed c_L data. Assuming that the healthy dentin will always exhibit a measured minimum density (ρ) that varies between 1.95 and 2.05 kg/dm^3 and that its Poisson's ratio (ν_0) varies between 0.275 and 0.285, one can compute present values for the c_L-based Young's modulus that range from 12 GPa to about 26.5 GPa.

	Elastic behavior (ρ=2.00±0.05 kg/dm^3; ν_0=0.28±0.005)		
	Velocity (20 MHz) c_L (m/s)	Wave number k	Young's modulus E (GPa)
Maximum, LIT[40,41]	4000+40=4040	31 103	24.6,…,26.5
LIT[41,19]*a*	4000−40=3960	31 718	23.6,…,25.4
LIT[42]*a*	3800+15=3815	32 926	21.9,…,23.6
LIT[19]*b*	3630	34 616	19.9,…,21.4
LIT[42]*b*	3000	41 886	13.5,…,14.6
…	…	…	…
Minimum	2850+4=2854	44 009	12.3,…,13.2

show a broader spectrum of wave numbers. Since this is not the topic of this paper, wave numbers are only given for reference purpose here (Table V).

B. Validation of the speed of sound

The literature values given for the speed of sound indicate considerable variability in sonic velocities for both, enamel and dentin.[9,18,19,21,23–26,38–42] The measured velocities of tangentially propagating ultrasonic waves in the sound dentin of the present human teeth have often been described in literature. One early article dealing with ultrasonic pulses in hard dental tissues already delivers an ultrasonic velocity in dentin of 4000 m/s.[41] Although this value is rather high (near the upper end of our experimental results), the probability that any tooth specimen of the present pool exhibits this value anywhere through its thickness and that this tangential velocity can be tracked in that specimen at least once (or no matter how often) is given by $p \geqslant 0.8$. That value is called the probability-of-detection (POD). The validated POD value is calculated for the utilized coplanar ground sections of the present study and may be even higher for completely evaluated dentin cores and/or any thinner longitudinal ground sections. Analogously, a reported[42] ultrasonic velocity of 3800 m/s (n=5) and 3630 m/s (n=4) in human and bovine dentin, respectively, is detected in every single specimen of the present study, thereby yielding a POD of 100%. Two additional horizontal cross sections are prepared for the purpose of comparing the main results of the c_L distribution with further velocity data of ultrasonic waves that propagate in nontangential tooth directions (axial direction).[19] The cross sections are also prepared to give a definite statement about the role of the tubules direction.

The values of Al-Nawas *et al.*[38] are not in accordance with the present findings. In contrast, Mezawa *et al.*[19] provide values between 3000 and 4000 m/s for the ultrasound velocity in the dentin. These values match our results. Their number of specimen—ten molars and ten premolars—is similar to our total of 21 teeth.

C. The role of tubule orientation

In contrast to most data derived from the longitudinal sections, the "lying v"-shaped region between *B* and *P* (Fig. 5) is partially representing the horizontal cross-sections, whose radial contour is distinctly associated with X=0.5. This observation has to be interpreted with regard to both (a) the principally distinct angle of incidence of the ultrasonic beam relative to the tubule direction and the tooth axis[36] (Fig. 6), and (b) the narrowed anatomical width at the L_5 plane found in most teeth (as compared to their maximal tooth widths; i.e., a relative width of $X > 0.45$ is seldom, and therefore there are fewer measured *p*-wave velocities assignable).

D. Relevance of the distribution of extreme *p*-wave velocities

The importance of the spatial distributions of the minimal velocities can be evaluated according to the manifestations of the underlying mechanical behavior of radiographically sound regions that probably had not been affected by any demineralization process. The relevance regarding a possible linking of velocity distributions with elasticity or correlations with the tubule density and the tubule orientation is obvious because any functional mapping of the minimal velocities and the maximal velocities allows for an economic but detailed inspection of a reduced number of selected regions-of-interest by means of acoustic microscopes or confocal laser scanning microscopes, respectively.[43–45] Researchers that account for the present appearance of the spatial distribution of minimal and maximal *p*-wave velocities can more easily decide whether the accuracy and reliability of any measurements can be improved for certain locations and where it is advantageous to take even thinner specimens. With this *a priori* knowledge it is simpler to choose between a unified thickness and a variable thickness. The most appropriate thickness of specimens can now be better estimated depending on their origin inside human teeth. These predictions can be advantageously used in any study that deals with mapping the Young's modulus[46] or the simultaneous measurement of sound velocity and thickness of extremely thin samples.[47] It is worthwhile to mention a prospective aspect because it helps to avoid some routine interpretations of artifacts that can be attributed to unwanted multiple reflections. According to this argument the spatial distributions of the minimal velocities can be used to derive and propose a clinically relevant threshold criterion for the desired maximal time-of-flights to be expected during any *in vivo* experiments. Useless signal portions that are erroneously sampled in the time domain beyond the limit of the threshold criteria would be attributed to multiple reflections and can therefore be partially omitted provided that the path lengths are known.[3] An additional outlook is given here for the purpose of how to depict further details of the underlying structures as well as to explain the necessity and importance of the spatial distribution of the maximal LUVs. Each time when the velocity of *p* waves occasionally exhibits any higher value than expected, the requirement for the minimal axial distance that should be resolved in that specific part of the

experiment is not fulfilled. Since the axial resolution depends also on the given operating frequency, the knowledge of the distribution of the maxima is of vital importance. It helps to avoid unresolved details or artifacts as well as to preset the proper operating frequencies of suitable transducers. This includes any desired inspection of the concentric multi-phase layers of the peritubular dentin and one may think of potential applications of the electromagnetic-resonance-ultrasound microscopy[46] and the time-resolved acoustic microscopy.[47] Furthermore, there are reports of early applications of acoustic microscopy to dental materials.[48,49]

E. Derived elasticity

The mentioned values of the present c_L data can be utilized for calculation of the elastic behavior. A combination of Eqs. (2), (4), (6), and (7) may deliver maximal elasticities for the given set of adjustable parameter coefficients and an assumed variation in density[5,20,50] and Poisson's ratio.[7,37,51–53] Analogously, a combination of Eqs. (2), (4), (6), and (8) [or (9)] gives mean (or individual) distributions of elasticity, provided that a continual range of elasticity is generated due to meaningful input quantities. The values for the Young's modulus derived from the present c_L data vary at least from 12.3 to 26.5 GPa according to Table V. Tyldesley[54] denotes the variation in the elastic modulus of $n=21$ female sound anterior specimens by 12.3 ± 0.34 GPa. Please note that this mean value of 12.3 GPa equals the present lower limit of the dentin's elasticity (as evaluated under the conditions set for minimum material density[5,20] and an average Poisson′s ratio[7,37,51–53] according to Table V), and the number of specimens is identical, too. Our estimated maximum value of 26.5 GPa is in accordance with a maximum elastic modulus of 25.4 GPa reported for longitudinal sections in a work of Miyazaki *et al.*[39] and lies in between the values reported for the stiffest direction in hydrated dentin (25.1 GPa) and the dried dentin (28.1 GPa).[37]

F. Padé shape functions and distinct approximations

In the application of Padé methods to signal processing a basic problem is to take into account the effect of measurement noise with respect to the computed coefficients.[33] Noise in this context may affect the relative width variable and the time jitter error during the determinations of the carrier frequency f_c and the time-of-flights (TOF) Δt_1 and Δt_2.

The present TOF measurements are considered for the economic use of a new MEES approach including macro-elements. Therefore, some linear fitting is performed by means of special interpolating functions. The choice of different approximations for different experimental data depends on the list of requirements a potential quantification has to fulfill. Considering Padé shape functions, RPs are utilized here to definitely cover the complete range of the present lowest *p*-wave velocities and to give a basis for any possible future upgrades that might account for even more extreme distributions. This means that the RP provided is "upgradable" as well as "release compatible" to adaptive simulations that utilize finite elements and a certain group of macro-elements.[32] Furthermore, the presented RPs and Table II can be quite useful for the purpose of excluding individual Poisson's ratios that might not be observed, reporting any imaginary parts of complex Poisson's ratios, as well as for modeling attempts that account for a pronounced anisotropy in the Poisson's ratio of wet dentin.[37] In addition to using the exact relationship between *p*-wave velocity and unidirectional Young's modulus, it is considered advantageous to utilize expanded alternative formulas as inputs for a new concept involving a continuum theory of microstructured brittle materials. Since any new concept is not the topic of the present paper, it should only be mentioned here that it may be desired to extensively correlate anisotropic Poisson's ratios, material densities, and the calculated velocities to the elastic behavior of such natural, functionally graded materials (*n*-FGMs) featuring gradual transition in composition, microstructure, and physical properties. The choice of different approximations for different experimental data includes the mean horizontal distribution of c_L (L_5 means), which is provided in a form where one term has the good fortune to explicitly show the average minimum velocity (2841 m/s) and where a second term contains the most frequently occurring *p*-wave velocity determined for one kind of teeth (3785 m/s). The choice of further approximations for different experimental data include the maximum curve and a general formula that makes it possible to fit all measured individual data with acceptable errors. It is obvious that the number of significant digits of coefficients is important for the quality of the fits. In this context, it is a surprisingly convincing observation that the descriptive power of the group of functions provided for the distribution of maximal *p*-wave velocities at a distance of 5 mm from the crown reference causes a necessity of no more than four significant digits. The hypothesis is formulated that this is true for at least one further level of inspection.[30] With respect to all possible levels of inspection, it should be mentioned that the overall applicability of this (and probably any) type of fitting is very doubtful in biological tissues that show similar degrees of complexity. In order to achieve no greater deviations than the errors described (as well as to account for quantitative results that are mathematically stable within an additional offset range of ±1.4 m/s), it should be stated here that the precision of the coefficients required to describe all of the individual data by means of the proposed general formula yields two coefficients with the necessity of four to five significant digits. This may be compared with the number of digits used for reporting the density distribution of enamel by means of data with up to five digits in recent publications,[20] while a typical table of standard reference data on pure substances, like water, contains 85 coefficients with an average number of 11.8 ± 3.8 digits.[55] The number of digits used for reporting the density distribution of cementum and dentin is $\leqslant4$ digits.[5,20,50] Increasing the absolute accuracy of the pulse-echo method on a daily reference basis, the repeatability of measured c_L data for non-newly immersed specimens is in between the exact values of the intrinsic accuracies of the pulse-echo overlap method and the pulse-superposition method.[56,57] The insertion of erroneous measurement "noise" as inputs for Padé shape approximations is assumed to initially contribute an amount of $\leqslant0.3\%$. The repeatability of c_L

data for newly immersed specimens is less (by one order of magnitude)—mainly due to slightly different paths, dynamic processes inside the dentinal tubules, and ray perturbation effects at distinct tubule directions. Since the variation of the individually fitted spatial distributions of the *p*-wave velocity in dentin reveals a great inhomogeneity from tooth to tooth, the overall applicability of these fittings is predominantly addressing the way how one can "homogenize" the local/spatial variations in longitudinal velocities. The way an effective *p*-wave velocity is composed and what kind of subregional velocities are involved is very important for any progress in the clinical use of ultrasound applied *in vivo* to hard dental tissues.

V. CONCLUSIONS

(1) Tangential *p*-wave velocities in human teeth are systematically measured at normal incidence and with inclined direction of wave propagation relative to the dentinal tubules. A least-squares error fitting is performed to quantitatively describe all discovered lowest/medium/maximal velocities and to illustrate the spatial variation as a function of relative tooth width locations. Several axial *p*-wave velocities (partially parallel and perpendicular to tubules) are determined, too. The lateral distribution of c_L is quantified for layers located at a 5-mm distance from the crown reference by using longitudinal sections of extracted human teeth. Horizontal scan lines are compared by standardized relative tooth width portions along a reduced distance in dentin.

(2) The lateral distribution of the ultrasonic velocities in longitudinal sections briefly increases with radial distances (i.e., increasing relative widths). The relative width is introduced as a running variable with which it is possible to unify all of the individually varying tooth widths and to compensate for a different number of roots.

(3) From radial direction data of perpendicularly irradiated ground sections of human dentin it is concluded that the spatial distribution of the c_{Lmin} can be described by means of a fractional polynomial law and will be helpful to avoid the diagnostic sampling of ultrasonic artifacts.

(4) The calculated values for the Young's modulus derived from the present c_L data vary from 12.3 to 26.5 GPa.

(5) The lateral distribution of tangential *p*-wave velocity in horizontal layers of human teeth at a corono-apical height level of 5 mm can be fully described by means of error evaluations and individual horizontal formulas of the form $c_L = a + bXX \ln(X)$.

(6) The amount of absolute residuals that deviate from the measured data by a given mean limit <100 m/s reaches a value of 100%.

(7) The spatial distribution of c_L is further characterized by its maximal envelope curve, which follows an empirical formula with three rounded integer values and may serve for selecting the best diagnostic operating frequency.

(8) The radial enlargement of the conical tubules towards the pulp plays a key role for the propagation characteristics of *p* waves. A moderate or high peripheral *p*-wave velocity of 3800 m/s is determined for a location with a quite extreme relative tubule area of 6.6% and a minor variation in tubule diameters of about 1 ± 0.2 μm.

ACKNOWLEDGMENTS

The author would like to thank Prof. Dr. Claus Löst (Medical Director) and Prof. Dr.-Ing. U. Faust (Emeritus and former Director of the Biomedical Engineering Institute). The author is grateful to Dr.-Ing. K.-M. Irion and Dipl.-Ing. Werner Nüssle for initiating related research in dental periodontology and for providing numerous data. I would also like to thank Dr. Birgit Linder-Lais, Dr. Michael Aschenbrenner, Dr. Angela Vogel, and Prof. Dr. Michael Brecx for providing numerous samples of extracted human teeth. Some aspects of the present material are linked to work supported, in part, by Grant No. Lo328/2-1,2 from the German Science Foundation (DFG).

APPENDIX A: LATERAL VARIATION OF THE CORONO-APICAL DECREASE

1. Locations

This evaluating procedure is performed at normative relative tooth width portions that range from $0.64w$ to $0.76w$ in specimens with a single root. A set of substitutes with relative tooth width portions laying within a window of relative x locations is defined by $0.826w \pm 0.06w$ for cases where specimens with two or more roots are involved [i.e., where the corresponding absolute positions $p(x,y)$ would otherwise represent the soft tissue inside the specimen's pulp chamber].

2. Procedures and details

For evaluation of the lateral variation of the corono-apical decrease, an arbitrary set of 15 corono-apical planes is utilized. Each plane is investigated as a function of normative relative tooth width portions regarding the c_L distribution in a single specimen.

3. Economics

No fitting is applied to the original data. The data range exhibiting a continual decrease may be of special interest.

APPENDIX B: HORIZONTAL FORMULA AND LATERAL ENVELOPE CURVES (L_5)

1. Locations

This evaluation is performed at all possible standardized relative tooth widths in between the full range from $X=0$ to $X=0.5$. Any relative width is not possible, if the corresponding absolute value represents the embedding resin or any air gap. This happens for any $a \leqslant 1.0$ and $x=aw$ if the individual width of the specimen at L_5 is already smaller than the individual maximal tooth width of that same specimen. The mentioned air gaps probably occur when the shrinkage of the embedding resin predominates its retention to the surface of the specimen during the curing.

2. Procedures and details

For an evaluation of the lateral distribution in the L_5 plane, the measured c_L values of every single specimen are processed by mathematical algorithms. A total of over 98 000 different equations are evaluated with respect to best fitting results. Five procedures are used, out of which procedures II and V are considered the most desired ones.

(1) Procedure I: Initial calculation of the mean c_L at a certain location and the next location utilizing data of all specimens; then the mean data are plotted as a function of location without fitting. A fitting is postprocessed for establishing the best Padé shape function with ten parameters that fits with an acceptable fitting quality (Fit.Std.Err.=10.4).
(2) Procedure II: Simultaneous fitting of all c_L as a function of location. From all of the generally best least squares fitting equations, the best fitting equation that comprises only three adjustable parameters is chosen.
(3) Procedure III: Fitting of the data of two subsets: (a) lowest $c_{L(L_5)}$min at 26 relative width locations and (b) highest $c_{L(L_5)}$max at 26 locations. From all of the generally best least squares fitting equations, (a) the best Padé shape function fits the first subset (no. 1) with an acceptable fitting quality and (b) the best alternative formula adequately fits the second subset (no. 2) by utilizing a required integer triplet of adjustable parameters with an acceptable fitting quality. The acceptable fitting quality is numerically defined by the following equation: Fit.Std.Err.$<(0.006/(3-\text{no.}))$
(4) Procedure IV: Fitting of the data like in procedure II but with an additional subset of data that came from the two horizontal sections. From all of the generally best least squares fitting equations, an arbitrary fitting equation that comprises five adjustable parameters is chosen. The correlation coefficient should follow the requirement $R^2>0.994$.
(5) Procedure V: Fitting of the individual c_L data as a function of relative width X. From all of the generally best least squares fitting equations, the best fitting equation that comprises only two adjustable parameters is chosen. A prerequisite is formulated that this type of fitting equation has to be appropriate for each of the individual specimens.

3. Individual residual deviation

The individual residual deviation (IRD) at every measured location inside human dentin is calculated from the difference between the actual velocity and the approximate value of the predicted velocity. Two threshold level criteria are defined that make an evaluated fitting equation incongruous and incompetent for further approval: via regression coefficients and relative residual errors (RREs). Dividing the IRD by the corresponding actual velocity at each location, the relative residual errors at all locations are calculated from these quotients. Furthermore, the procedure's fitting quality is determined by means of the regression coefficient (R^2). The fittings of procedure V are accepted (a) provided that RRE$<$ca.4% (by average, it is also exceptionally accepted that each specimen may solely show a single data point deviation that is greater than 4% but smaller than 14%) and (b) $R>0.79$ if $\mathcal{N}\leqslant 6$, or $R>0.86$ if $\mathcal{N}\leqslant 3$, or for $\geqslant$50% of all $\mathcal{N}'=21$ specimens any $R\geqslant 0.92$ if $\mathcal{N}=1$ ($R^2>0.84$). Here, $\mathcal{N}$ is the number of individual specimens that are commonly evaluated.

4. Economics

With respect to economical aspects, it is desirable to summarize the individual relative residual errors by a mean value and a calculated standard deviation, too. This is helpful, provided that all of these individual mean RREs are smaller than one tenth of the full dentin's range of c_L values. Since the latter quantity corresponds to about 2400 m/s,[9] it is considered desirable to find a fitting equation that accounts for a "mean |Residual|" that is smaller than 240 m/s.

APPENDIX C: FUNCTION OF PRIORITY

A new function of priority is considered here to be more appropriate than all the competing ones tested before, provided that we have the following.

(1) The newly evaluated type of function provides a better fit of the measured data for the horizontal scans, and this requirement is fulfilled for all specimens. The quality of the fit is quantitatively evaluated by means of five criteria:
  (1.1) the number of absolute residuals that deviate from the measured data by a given limit of, e.g., less than 100 m/s;
  (1.2) the percentage of those specimens that can be approximated by the function in such a manner that the calculated mean ±SD of the dentin's absolute residuals exceeds neither a value of 2.2% (3.3%) relative to the measured maximal (minimal) LUV nor 100±100 m/s;
  (1.3) R^2;
  (1.4) the number of negative relative residuals that exceed a given limit of $-(4.0\%-0.1\%)=-3.9\%$; and
  (1.5) the number of positive relative residuals that exceed a given limit of +3.9%.
(2) The newly evaluated type of function makes it possible to quantitatively compare all horizontal scan lines without losing accuracy when using a reduced number of coefficients (and no increase in digits required).
(3) The newly evaluated type of function is suitable to approximate the desired fitting results in a significantly shorter computation time.
(4) The newly evaluated type of function is providing a better link to any important formulas and any descriptive work of other studies that deal with parameter patterns and related physical properties.
(5) The newly evaluated function is superior due to the reduced number of digits (figures) that have to be kept for each of its coefficients while maintaining the required precision.

(6) The calculated coefficients of the newly evaluated function are more convincing because they can uniformly be utilized for all the specimens in the same manner.
(7) The set of coefficients of the newly evaluated function makes it possible to integrate data in an efficient and more meaningful manner into models designed for medical element engineering and simulation (MEES).

APPENDIX D: HIERARCHY OF CRITERIA

Although criteria (6) may be the "gold standard," the weighing factors may also be sorted in a hierarchical manner that excludes (6). With increasing powers of modern computers, (3) becomes also less important and an alternative process is less cumbersome. It is more effective to start with repeated evaluations according to (1.3), thereby successively improving (2) and (5) for the purpose of (4) and (7) by iteratively choosing stronger cutoff limits for (1.1), (1.4), and (1.5) until (1.2) reaches a 100% success rate.

[1]H. E. Schroeder, *Orale Strukturbiologie* (Thieme Verlag, Stuttgart, 1987), Chap. 1.1.6—1.6, pp. 20–186.
[2]N. P. Chandler, T. R. Pitt Ford, and B. D. Monteith, "Coronal pulp size in molars: a study of bitewing radiographs," Int. Endod. J. **36**, 757–763 (2003).
[3]C. John, "Directing ultrasound at the cemento-enamel junction (CEJ) of human teeth: I) Asymmetry of ultrasonic path lengths," Ultrasonics **43**, 467–479 (2005).
[4]M. Garala, S. Kuttler, P. Hardigan, R. Steiner-Carmi, and S. Dorn, "A comparison of the minimum canal wall thickness remaining following preparation using two nickel-titanium rotary systems," Int. Endod. J. **36**, 636–642 (2003).
[5]G. V. Black, "An investigation of the physical characters of the human teeth in relation to their diseases, and to practical dental operations, together with the physical characters of filling-materials," Dent. Cosmos **37**, 353–421 (1895).
[6]A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).
[7]H. H. K. Xu, D. T. Smith, S. Jahanmir, E. Romberg, J. R. Kelly, V. P. Thompson, and E. D. Rekow, "Indentation damage and mechanical properties of human enamel and dentin," J. Dent. Res. **77**, 472–480 (1998).
[8]J. L. Katz, "Anisotropy of Young's modulus of bone," Nature (London) **283**, 106–107 (1980).
[9]C. John, "The corono-apically varying ultrasonic velocity in human hard dental tissues," J. Acoust. Soc. Am. **116**, 545–556 (2004).
[10]H. S. Yoon and J. L. Katz, "Temperature dependence of the ultrasonic velocities in bone," IEEE Proc. Ultrasonics Symp. **CH1482-9/79/0000-0395** (1979), pp. 395–398.
[11]A. Kishen, U. Ramamurty, and A. Asundi, "Experimental studies on the nature of property gradients in the human dentine," J. Biomed. Mater. Res. **51**, 650–659 (2000).
[12]H. Hülsmann, M. Heckendorff, and Á. Lennon, "Chelating agents in root canal treatment: mode of action and indications for their use," Int. Endod. J. **36**, 810–830 (2003).
[13]K. Verdelis, M. A. Crenshaw, E. P. Paschalis, S. Doty, E. Atti, and A. L. Boskey, "Spectroscopic imaging of mineral maturation in bovine dentin," J. Dent. Res. **82**, 697–702 (2003).
[14]D. H. Pashley, "Clinical correlations of dentin structure and function," J. Prosthet. Dent. **66**, 777–781 (1991).
[15]L. Vasiliadis, A. Darling, and B. Levers, "The histology of sclerotic human root dentine," Arch. Oral Biol. **28**, 693–700 (1983).
[16]I. A. Mjör, "Age changes in the teeth," in *Geriatric Dentistry*, edited by P. Holm-Pedersen and H. Löe (Munksgaard, Göteborg, Sweden, 1986), pp. 94–101.
[17]M. G. Duncanson and E. Korostoff, "Compressive viscoelastic properties of human dentin: I. Stress-relaxation behavior," J. Dent. Res. **54**, 1207–1212 (1975).
[18]C. Löst, K. M. Irion, C. John, and W. Nüssle, "Two-dimensional distribution of sound velocity in ground sections of dentin," Endod. Dent. Traumatol. **8**, 215–218 (1992).
[19]S. Mezawa, T. Kawato, K. Yoshida, H. Nozaki, T. Saito, K. Tamura, and M. Onozawa, "Evaluation of human tooth structure with the ultrasonic imaging technique," J. Oral Sci. **41**, 191–197 (1999).
[20]R. S. Manly, H. C. Hodge, and L. E. Ange, "Density and refractive index studies of dental hard tissues. II. Density distribution curves," J. Dent. Res. **18**, 203–211 (1939).
[21]S. Lees, "Specific acoustic impedance of enamel and dentine," Arch. Oral Biol. **13**, 1491–1500 (1968).
[22]J. E. Lynch and M. K. Hinders, "Ultrasonic device for measuring periodontal attachment levels," Rev. Sci. Instrum. **73**, 2686–2693 (2002).
[23]C. Löst, C. John, K. M. Irion, and W. Nüssle, "Dentincharakterisierung mittels zweidimensionaler Ultraschallgeschwindigkeitsprofile (Characterization of the dentin by a 2-dimensional ultrasonic velocity profile)," Schweiz Monatsschr Zahnmed **104**, 20–24 (1994) (in German).
[24]C. John and C. Löst, "Acoustic images of human teeth," in *Computer Assisted Radiology—Excerpta Medica*, Int Congr Ser No. 1124 (Elsevier Science, Amsterdam, 1996), pp. xxvii, 902–908, 1089.
[25]C. John, "Recovering echo signals for in-vitro characterization of hard dental tissues," in *Nondestructive Characterization Material VIII* (Plenum, New York, 1998), pp. 763–768.
[26]C. John, K. M. Irion, W. Nüssle, and C. Löst, "The resolution of a 2-dimensional ultrasonic velocity profile of human tooth sections," Schweiz Monatsschr Zahnmed **104**, 25–30 (1994) (in German).
[27]R. H. Eichler, "Apparatus for measurement of sound velocity in a workpiece," U.S. Patent No. 3,690,155 (1972).
[28]M. S. Hughes and D. K. Hsu, "An automated algorithm for simultaneously producing velocity and thickness images," Ultrasonics **32**, 31–37 (1994).
[29]D. K. Hsu and M. S. Hughes, "Simultaneous ultrasonic velocity and sample thickness measurement and application in composites," J. Acoust. Soc. Am. **92**, 669–675 (1992).
[30]C. John, "The laterally varying ultrasonic velocity in the dentin of human teeth," J. Biomech. (in press) http://dx.doi.org/10.1016/j.jbiomech.2005.08.004.
[31]M. E. Alarney, C. B. Phillips, and D. Song, "Form difference computations in restorative dentistry utilizing the macroelement method," J. Dent. Res. **75**, 1464–1476 (1996).
[32]M. E. Alarney, "The use of the boundary element method for biological morphometrics," J. Biomech. **28**, 609–616 (1995).
[33]P. Barone, "On the distribution of poles of Padé approximants to the Z-transform of complex Gaussian white noise," J. Approx. Theory **132**, 224–240 (2005).
[34]J. H. Kinney, M. Balooch, G. W. Marshall, and S. J. Marshall, "A micromechanics model of the elastic properties of human dentine," Arch. Oral Biol. **44**, 813–822 (1999).
[35]M. Balooch, I.-C. Wu-Magidi, A. Balazs, A. S. Lundkvist, S. J. Marshall, G. W. Marshall, W. J. Siekhaus, and J. H. Kinney, "Viscoelastic properties of demineralized human dentin measured in water with atomic force microscope (AFM)-based indentation," J. Biomed. Mater. Res. **40**, 539–544 (1998).
[36]L. G. Watanabe, G. W. Marshall, Jr., and S. J. Marshall, "Dentin shear strength: Effects of tubule orientation and intratooth location," Dent. Mater. **12**, 109–115 (1996).
[37]J. H. Kinney, J. R. Gladden, G. W. Marshall, S. J. Marshall, J. H. So, and J. D. Maynard, "Resonant ultrasound spectroscopy measurements of the elastic constants of human dentin," J. Biomech. **37**, 437–441 (2004).
[38]B. Al-Nawas, K. A. Grötz, E. Rose, H. Duschner, P. Kann, and W. Wagner, "Using ultrasound transmission velocity to analyze the mechanical properties of teeth after in vitro, in situ, and in vivo irradiation," Clin. Oral Investig. **4**, 168–172 (2000).
[39]M. Miyazaki, H. Inage, and H. Onose, "Use of an ultrasonic device for the determination of elastic modulus of dentin," J. Oral Sci. **44**, 19–26 (2002).
[40]S. Y. Ng, P. A. Payne, N. A. Cartledge, and M. W. J. Ferguson, "Determination of ultrasonic velocity in human enamel and dentine," Arch. Oral Biol. **34**, 341–345 (1989).
[41]G. Kossoff and C. J. Sharpe, "Examination of the contents of the pulp cavity in teeth," Ultrasonics **4**, 77–83 (1966).
[42]F. E. Barber, S. Lees, and R. R. Lobene, "Ultrasonic pulse-echo measurements in teeth," Arch. Oral Biol. **14**, 745–760 (1969).
[43]T. Kundu and C. S. Jorgensen, "Measuring elastic properties of bones and silicon from V(z) curve generated by multiply reflected signals," Ultrasonics **39**, 515–524 (2002).
[44]R. G. Maev, L. A. Denisova, E. Y. Maeva, and A. A. Denissov, "New data on histology and physico-mechanical properties of human tooth tissue obtained with acoustic microscopy," Ultrasound Med. Biol. **28**, 131–136

(2003).
[45]C. S. Jorgensen and T. Kundu, "Measurement of material elastic constants and density of cancellous bone: a micromechanical analytic study using a 1 GHz acoustic microscope," J. Orthop. Res. **20**, 151–158 (2002).
[46]J. Tian, H. Ogi, T. Tada, and M. Hirao, "Young's modulus mapping on SCS-6 SiC_f_/Ti–6Al–4V composite by electromagnetic-resonance-ultrasound microscopy," J. Appl. Phys. **94**, 6472–6476 (2003).
[47]V. Hänel, "Measurement of sound velocity and thickness of thin samples by time-resolved acoustic microscopy," J. Appl. Phys. **84**, 668–670 (1998).
[48]J.-I. Kushibiki, K. L. Ha, H. Kato, N. Chubachi, and F. Dunn, "Application of acoustic microscopy to dental material characterization," IEEE Conference Proceedings 0090–5607/87/0000–0837 (1987), pp. 837–842.
[49]S. D. Peck and G. A. D. Briggs, "The caries lesion under the scanning acoustic microscope," Adv. Dent. Res. **4**, 168–172 (2000).
[50]C. John, "The density of human teeth after immersion in saline water," J. Dent. Res., **77** (Supp.), 847 (1998).
[51]R. S. Gilmore, R. P. Pollak, and J. L. Katz, "Elastic properties of bovine dentine and enamel," Arch. Oral Biol. **15**, 787–796 (1969).
[52]A. Verluis, D. Tantbirojn, and W. H. Douglas, "Do dental composites always shrink toward the light?" J. Dent. Res. **77**, 1435–1445 (1998).
[53]C. E. Renson and M. Braden, "The experimental deformation of human dentine by indenters," Arch. Oral Biol. **16**, 563–572 (1971).
[54]W. R. Tyldesley, "The mechanical properties of human enamel and dentine," Br. Dent. J. **106**, 269–278 (1959).
[55]GSSSD 98–86, "Water Specific volume and enthalpy at temperatures 0–800 °C and at pressures 0.001–1000 MPa," in *Tables of Standard Reference Data* (Standards Press, Moscow, 1984).
[56]E. P. Papadakis, "Absolute accuracy of the pulse-echo overlap method," J. Acoust. Soc. Am. **52**, 843–846 (1972).
[57]K. J. Parker, L. S. Taylor, S. Gracewski, and D. J. Rubens, "A unified view of imaging the elastic properties of tissue by superposition method," J. Acoust. Soc. Am. **117**, 2705–2712 (2005).

Nonoptimal propagation of advertisement calls of midwife toads in Iberian habitats

Mario Penna[a)]
Programa de Fisiología y Biofísica, Facultad de Medicina, Universidad de Chile, Casilla 70005, Correo 7, Santiago, Chile

Rafael Márquez and Jaime Bosch
Fonoteca Zoológica, Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales (CSIC), José Gutiérrez Abascal 2, 28006 Madrid, Spain

Eduardo G. Crespo
Departamento de Biologia Animal, Centro de Biologia Ambiental da Faculdade de Ciencias da Universidade de Lisboa, Bloco C2, Campo Grande.1749-016 Lisboa, Portugal

(Received 5 April 2005; revised 13 October 2005; accepted 14 October 2005)

This study compares the efficiency of transmission of the advertisement calls of two species of midwife toads, *Alytes cisternasii* and *A. obstetricans*, in both native and non-native habitats in the Iberian Peninsula. Recorded calls of both species and pure tones were broadcast at ten sites native to either the relatively small *A. cisternasii* or the larger *A. obstetricans*. A large variation in the patterns of excess attenuation between localities was observed for calls measured at distances of 0.5 to 8 m from a loudspeaker. However, attenuation rates were higher for calls of both species in habitats of *A. obstetricans* relative to habitats of *A. cisternasii*. The calls of *A. obstetricans* experienced lower attenuation rates than those of *A. cisternasii* in both conspecific and heterospecific localities. Thus, although *A. cisternasii* occupies habitats more favorable for sound transmission, its advertisement call spectrum is not optimized for these habitats; the calls of *A. obstetricans* suffer less attenuation in *A. cisternasii* habitats. This result argues against the notion that spectral features of the calls are adapted to enhance transmission efficiency in natural habitats, and suggests that differences in call dominant frequency between the two species result from constraints imposed by selection on body size. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2149769]

PACS number(s): 43.80.Gx, 43.80.Ev, 43.80.Ka [JAS] Pages: 1227–1237

I. INTRODUCTION

Signals conveying information to distant receivers are expected to evolve features that maximize their propagation in the native habitats of the emitters. Optimal relationships of this kind have been shown to occur for acoustic signals of birds, such that songs of species dwelling in different environments possess temporal and spectral features that contribute to improve their transmission in native habitats. Birds inhabiting forests produce songs having less amplitude modulations and lower frequencies relative to species from open environments (e.g., Morton, 1975; Gish and Morton, 1981; Shy, 1983; Ryan and Brenowitz, 1985; Sorjonen, 1986; Slabbekoorn and Smith, 2002).

The advertisement calls of male anurans attract mates and provide cues for phonotactically behaving females to discriminate among potential mates within a chorusing aggregation (e.g., Gerhardt and Huber, 2002). Few studies have explored associations between habitat and call structure in anurans, and no optimization such as reported for birds has been found to occur in these vertebrates. In a comparative study of tropical South American anurans grouped by their native environments, broadly classified as open and forested, Zimmerman (1983) found that features of vocalizations were dependent on phylogeny and not on the species habitat. Experimental studies examining the properties of diverse habitats for transmission of broadcast calls of several anuran species have reported a lack of relationship between call structure and habitat, both in South American temperate (Penna and Solís, 1998) and Central American tropical (Kime *et al.*, 2000) environments.

However, in two subspecies of the North American frog *Acris crepitans*, there is some evidence on environmental selection on call structure. Males from forested habitats produce calls having higher frequencies and faster rates that are not subject to degradation in these environments to the same extent as calls of populations from open habitats which have slower rates and lower frequencies (Ryan *et al.*, 1990).

Midwife toads *A. cisternasii* and *A. obstetricans* are primitive anurans widely distributed in the Iberian peninsula. These toads produce simple calls consisting of tonal notes lacking frequency modulation. Considerable overlap in the call dominant frequency and duration occur between both species. The call of *A. obstetricans* has a lower dominant frequency (about 1100 Hz) relative to the call of *A. cisternasii* (about 1400 Hz; see Fig. 1) and this difference is consistent with differences in body size between species, *A. ob-*

[a)]Electronic mail: mpenna@med.uchile.cl

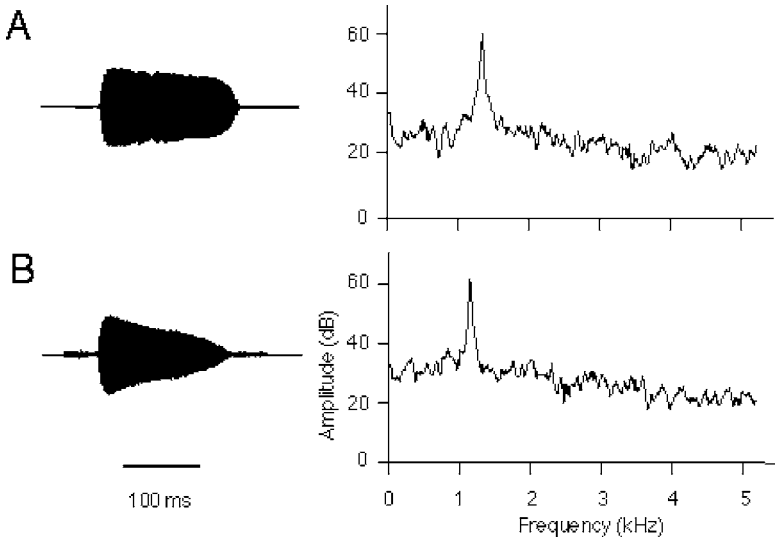


FIG. 1. Oscillograms and power spectra of representative advertisement calls used for sound propagation experiments: (a) *Alytes cisternasii* and (b) *A. obstetricans* Air temperatures during recordings of these calls were 14 °C and 12 °C, respectively.

stetricans being slightly larger than *A. cisternasii* (Márquez and Bosch, 1995).

In the central region of the Iberian Peninsula, *A. obstetricans* is distributed in mountain ranges at elevated sites. In the rest of its distribution, in the north and west of the Iberian Peninsula and western Europe, this toad occurs in a wide variety of habitats: alpine grasslands, deciduous and coniferous forests, and lowlands with grass or bush vegetation cover. The breeding sites of this toad are associated with permanent bodies of water and throughout most of its distribution, mating takes place during a relatively extended period, from late spring to late summer (Crespo, 1982; Reading and Clarke, 1988; Márquez, 1992; Böll and Linsenmair, 1998). *A. cisternasii* occurs in more xeric environments in the southwest quarter of the Iberian Peninsula, in Mediterranean oak savannas where live oaks (*Quercus ilex* and *Q. suber*) are the predominating trees. This toad breeds near temporary creeks and pools, the breeding period in the southern Iberian Peninsula being considerably shorter than the one of *A. obstetricans*. In these latitudes, breeding is usually completed within 2 or 3 weeks following the first heavy autumn rains in September or October (Crespo, 1982; Rodríguez-Jiménez, 1984; Márquez, 1992; Márquez and Crespo, 2002). The breeding sites of the two species have similar topographies, consisting of open spaces with rocky and sandy substrates and low herbaceous vegetation. Males of both species call and mate on land, and carry the egg clutches attached to their legs for about 1 month, until they release them at the shores of bodies of water. Both species occur in sympatry in Central Spain, and in this area preferences for dominant frequencies prevent heterospecific matings (Márquez and Bosch, 1997b). During the breeding activity of these toads, no other anurans call from the ground surface, the sound environment in which the two *Alytes* species communicate being relatively simple, devoid of substantial acoustic interference from biotic sources.

Females of both species show a tendency to orient phonotactically towards low-frequency calls within the range of conspecific males. These preferences are congruent with the mating success of larger males, indicating that the spectral structure of the advertisement calls is under intersexual selection pressure favoring lower dominant frequencies (Márquez, 1995a, b; Márquez and Bosch, 1997a). Males also call more actively in response to synthetic stimuli having lower dominant frequencies (Bosch and Márquez, 1996), indicating that this trait is important in acoustic competition and probably subject to intrasexual selection as well.

In addition, the dominant frequency of the calls of *A. cisternasii* and *A. obstetricans* is a component with low intraindividual variability. Males do not alter their dominant frequency in contests with other males (Bosch and Márquez, 2001) or with synthetic stimuli (Bosch and Márquez, 1996), as has been reported to occur for males of other anuran species (López *et al.*, 1988; Wagner, 1989; Bee and Perril, 1996). In contrast with the dependence of spectra of the calls on individual characteristics, the temporal structure of the vocalizations is in agreement with the phylogenetic relationships of these toads. Specifically, the effect of temperature on call duration is related to the taxonomic status of Iberian populations of *Alytes* (Márquez and Bosch, 1995).

The simplicity of the ground acoustic environments where just *A. cisternasii* and *A. obstetricans* and no other anurans communicate, the individual invariance, and the lack of phylogenetic determination of the dominant frequencies of their calls render these toads amenable for the study of influences of environmental pressures on the structure of their long-range vocalizations. However, since males and females display preferential responses for calls of low dominant frequencies, sexual selection pressures acting upon acoustic signals are to be taken into account when assessing the environmental determinants of signal structure.

Here, we report on experiments of propagation of pure tones of different frequencies and advertisement calls of *A. cisternasii* and *A. obstetricans* in Iberian localities native to both species. We aim to determine how environmental selection pressures affect the spectral structure of the advertisement calls of these toads. The main question we address is whether habitats of *A. cisternasii* and *A. obstetricans* differ in their transmission properties, such that calls of each species are more efficiently propagated in their respective native habitats. Alternatively, habitats imposing larger constraints

TABLE I. Environmental conditions in the study sites of propagation of vocalizations of *A. cisternasii* and *A. obstetricans* in the Iberian Peninsula.

Locality	Latitude (N), Longitude (W)	Test date	Native species	Air temperature (°C)	Humidity (%)	Substrate
Grandola	38°06′,08°34′	January 2000	*A. cisternasii*	8–10	73	Sandy soil, grass, rock outcrops
Alcaçovas	38°06′,08°07′	January 2000	*A. cisternasii*	10–11	74	Sandy soil, grass, rock outcrops
Paço da Quinta	38°41′,07°44′	January 2000	*A. cisternasii*	14–15	53	Sandy soil, grass, rock outcrops
Laranjal	38°41′,07°46′	January 2000	*A. cisternasii*	7–9	72	Sandy soil, grass, rock outcrops
Almadén de la Plata	37°47′,06°04′	September 2002	*A. cisternasii*	24	65	Sandy soil, grass, rock outcrops
Puente del Duque	40°20′,05°10′	February 2000	*A. obstetricans*	12	51	Grass, rock outcrops
Peñalara	40°45′,04°00′	February 2000	*A. obstetricans*	2–6	62	Grass
Lago Enol	43°16′,05°00′	May 2001	*A. obstetricans*	20–26	60	Grass, rock outcrops
Covadonga	43°15′,04°59′	May 2001	*A. obstetricans*	35	58	Grass, rock outcrops
Lago de la Cueva	43°03′,06°06′	June 2001	*A. obstetricans*	17	72	Rubble, rock outcrops

for sound propagation may associate with calls that are better transmitted.

II. METHODS

A. Study sites

Sound propagation experiments were conducted in ten localities within the distribution ranges of *A. cisternasii* (five localities) and *A. obstetricans* (five localities) in the Iberian peninsula. Localities of *A. cisternasii* were Grandola, Alcaçovas, Paço da Quinta, and Laranjal in Central Portugal, and Almadén de la Plata in southern Spain. Localities of *A. obstetricans* were Puente del Duque, Peñalara, Lago Enol, Covadonga, and Lago de la Cueva in Central and Northern Spain. The geographical coordinates of the study sites are listed in Table I.

Because of constraints in traveling schedules, the tests at different localities were conducted at different times of the year. Measurements at four sites of *A. cisternasii* (Grandola, Alcaçovas, Paço da Quinta, and Laranjal) were conducted during winter in January of 2000, a time of the year when isolated individuals persisted calling in the breeding areas, and measurements in one site (Almadén de la Plata) were conducted during the peak of the breeding period in early autumn of 2002. Sound propagation experiments in two of the localities of *A. obstetricans* (Puente del Duque and Peñalara) were carried out during winter in February of 2000 and in Lago Enol, Covadonga, and Lago de la Cueva were conducted during late spring in May and June of 2001. All the test sites of *A. cisternasii* were in open forests of live oaks (*Quercus ilex* and *Q. suber*). The substrates over which measurements were conducted in winter were sandy soil with grass and rock outcrops in a patchy distribution. As for the habitats of *A. obstetricans*, the test site in Puente del Duque was in a forest of *Pinus sylvestris* and measurements were made over soil with green grass and stones in a patchy distribution. Peñalara is an alpine grassland devoid of trees, and tests were made over a relatively homogeneous surface of green grass. The test sites in Lago Enol and Covadonga were grass pastures with occasional rocky outcrops and sparse areas of eroded soil. In Lago de la Cueva measurements were conducted across a steep slope devoid of vegetation with abundant rocks and loose rubble on the shore of a mountain lake.

During the propagation experiments, the atmosphere was calm. Only slight gusts of wind occurred occasionally, during which measurements were suspended. Recordings were conducted in the afternoon and evening, time when the vocal activity of the toads begins. This avoided interference from the vocal activity of toads at night in the test areas during the breeding period. These measurements are assumed to apply to night hours during which frogs vocalize, and to breeding periods for cases in which measurements were conducted out of the breeding period, given that air temperature and humidity have negligible effects on sound propagation at the frequencies and distances from the source that we tested (Harris, 1966; Piercy and Daigle, 1991). Atmospheric conditions have been reported to affect attenuation of acoustic signals, but at distances beyond about 50 m (Van Staaden and Römer, 1997), well out of the 8-m range explored in our study. Measurements of tone propagation were conducted at a test site in different atmospheric conditions to test for the validity of the assumption of the lack of significant influences of atmospheric conditions on our field experiments (see below).

B. Playback calls

Advertisement calls of 12 *A. cisternasii* and 10 *A. obstetricans* were used for propagation experiments. The calls of these toads were recorded in two localities, and were chosen to encompass the range of variation of the dominant frequency of these signals in each species. The calls of the 12 males of *A. cisternasii* were recorded in Mérida (Badajoz, N 38° 58,W 6° 22') at temperatures of 12–22 °C. The calls of the 10 males of *A. obstetricans* were recorded in Formigal (Huesca, N 42° 47'W 0°23') at temperatures of 8–15 °C. The average dominant frequencies of the calls of *A. cisternasii* and *A. obstetricans* were 1473 Hz (range =1334–1590) and 1207 Hz (range=1050–1349), respectively. Six calls of each toad were edited with SOUNDMAKER 1.0.4 software on an Apple Macintosh 4100 laptop computer with 0.5-s intercall intervals and used in the propagation experiments. One call of each species is shown in Fig. 1.

The dominant frequencies of the playback calls are in partial agreement with the dominant frequencies of the calls of animals from the sites where the tests were conducted. Call dominant frequencies and body sizes for the populations

TABLE II. Size and dominant frequencies of the advertisement calls of populations of *Alytes* included in this study. Averages and ranges in parentheses.

Species and Locality	Snout-vent length(mm)	Dominant frequency(Hz)	Reference
A. cisternasii Mérida (Playback calls)	36.7 (27–42) $N=58$	1473 (1334–1599)[a], [d] $N=12$	Márquez and Bosch, 1995
A. cisternasii Southern Portugal	35.7 (32–39) $N=15$	1457 (1350–1670)[a] $N=39$	Crespo *et al.*, 1989
A. cisternasii Almadén de la Plata	35.2 (32–39) $N=12$	1472 (1292–1582)[a] $N=48$	R. Márquez and M. Penna, unpublished data
A. obstetricans Formigal (Playback calls)	42.6 (36–52) $N=24$	1207 (1050–1349)[b], [d] $N=10$	Márquez and Bosch, 1995
A. obstetricans Peñalara	44.9 (42–50) $N=16$	1159 (1035–1375)[b] $N=30$	Márquez and Bosch, 1995
A. obstetricans Puente del Duque	37.0 (32–42) $N=57$	1330 (1191–1520)[c] $N=57$	Márquez and Bosch, 1995
A. obstetricans Lago de la Cueva	34.9 (27–41) $N=8$	1317 (1148–1401)[c] $N=30$	R. Márquez, unpublished data

[a]Indicate populations having similar dominant frequencies (Tukey-Kramer test, $P>0.05$).
[b]Indicate populations having similar dominant frequencies (Tukey-Kramer test, $P>0.05$).
[c]Indicate populations having similar dominant frequencies (Tukey-Kramer test, $P>0.05$).
[d]Values listed for dominant frequency correspond to the calls used for playbacks.

of *A. cisternasii* and *A. obstetricans* included in this study drawn from the literature and unpublished data are listed in Table II. Data for the sites of Grandola, Alcaçovas, Paço da Quinta, and Laranjal in southern Portugal correspond to the study by Crespo *et al.* (1989), conducted in this region. No data are available for the populations of Lago Enol and Covadonga, but the values of the population of Lago de la Cueva, a few kilometers distant, are assumed to apply to these sites. A one-way ANOVA showed significant differences among the dominant frequencies of the calls of the seven populations analyzed: Mérida, southern Portugal, Almadén de la Plata, Formigal, Peñalara, Puente del Duque, and Lago de la Cueva ($F_{6,246}=98.19$, $P<0.0001$). *Post hoc* comparisons (Tukey-Kramer test, $P<0.05$) indicated that the dominant frequency of the calls of *A. cisternasii* of Mérida used for playbacks did not differ from the calls of this species of southern Portugal and Almadén de la Plata, but differed significantly from the three populations of *A. obstetricans*. Also, the dominant frequency of the calls of *A. obstetricans* of Formigal did not differ from the calls of Peñalara and differed significantly from the calls of the two populations of *A. cisternasii*. However, the dominant frequencies of the calls of *A. obstetricans* of Formigal and Peñalara were significantly lower than the calls of Puente del Duque and Lago de la Cueva. The dominant frequencies of these populations also were significantly lower than those of the populations of *A. cisternasii* included in the analysis. The average dominant frequency is significantly correlated with the average body size of these populations ($R^2=0.74$, $P<0.016$).

C. Experimental procedures

Twenty-seven pure tones of 0.2–5.0 kHz and 500-ms duration were generated with SOUNDMAKER 1.0.4 software on an Apple Macintosh 4100 laptop computer. The continuous tones and playback calls were amplified and delivered through a self-amplified loudspeaker (Phillips SBC BA130). The playback volume was adjusted at a level at which no distortion was observed for any of the frequencies used after recording the signals at 0.5 m from the loudspeaker and displaying them on the laptop computer (see below). The peak SPLs of the calls measured at 0.5 m in front of the loudspeaker were within the range of 94–102 dB peak SPL for the calls of the two species. The frequency response of the loudspeaker was within ±5 dB between 0.3 and 3.0 kHz and within ±8 dB between 0.3 and 5.0 kHz. The SPL of the 0.2-kHz tone was typically 10 dB below the levels measured at 0.3 kHz and at higher frequencies. The peak SPLs of the tones measured at 0.5 m in front of the loudspeaker were within the range of 84–102 dB peak SPL. The unequal frequency response of the loudspeaker did not affect measurements, since the synthetic and natural signals used were tonal, and attenuation was obtained by substracting amplitudes of signals at a given distance from amplitudes of the same signal at 0.5 m from the loudspeaker (see below).

In the study sites the loudspeaker was placed at positions typically occupied by calling males of *A. cisternasii* and *A. obstetricans*. Peak SPLs of the sounds were measured with the microphone of a sound-level meter (Brüel & Kjær 2230 or 2238) fitted with foam windshields and 10-m extension cables. The instruments were calibrated with a portable calibrator (Brüel & Kjær 4230 or 4231) before and after each recording session. The microphone was positioned at 0.5, 1, 2, 4, and 8 m in front of the loudspeaker. Peak SPLs were measured for continuous tones of 0.7, 0.8, 0.9, and 1.0 kHz at each distance in all study sites, to check for the accuracy of the calibration procedure for the rest of the recorded signals. The microphone elevation—corresponding to the radius

of the foam wind shield—was constant at 5 cm for all distances. Signals detected were fed via a connecting cable from the ac output of the sound-level meter into one channel of a digital tape recorder (Sony TC-D8 or Tascam DA-P1). The recording level of the tape recorder was kept constant for all measurements, and at distant locations from the loudspeaker the sensitivity of the sound-level meter was increased in the discrete steps provided by the instrument, in order to input detectable signals into the tape recorder.

Pure tones and playback calls were recorded at different distances and, after the measurements were completed, the 1-kHz tone of a Brüel & Kjaer calibrator was recorded with the sound-level meter microphone. The peak SPLs of the recorded signals were determined relative to the amplitude of the recorded calibration tone, which measured 96.8 dB peak SPL. Excess attenuation (attenuation in excess of that expected due to spherical spreading) of pure tones and playback calls was calculated for distances of 1, 2, 4, and 8 m from the loudspeaker, relative to measurements at 0.5 m. These distances were chosen to facilitate calculations of excess attenuation at distances doubling the preceding one and to allow comparisons with a previous study (Penna and Solís, 1998). Values predicted by spherical spreading were calculated with the equation: dB attenuation=20 log [far distance (m)/0.5(m)]. This value was subtracted from the difference between the values measured in dB SPL at 0.5 m from the loudspeaker and those measured at the corresponding far distances. Positive and negative values indicated that the sound attenuated at rates higher and lower, respectively, relative to SPLs predicted by spherical attenuation for each distance. Maximum rms and peak SPLs of ambient noise occurring during 15-s periods in the study sites previous to the sound propagation experiments were measured with the sound-level meter microphone placed on the transect along which natural and synthetic sounds were broadcast.

To test for the influence of atmospheric conditions on sound attenuation, propagation tests were conducted in a Mediterranean habitat in the locality of Los Maitenes in Central Chile (S 34° 40' W 71° 27'), on 1 October 2005. The tests were conducted in a steppe of *Acacia caven*, across a surface of short green grass. Four 3-s-duration pure tones of 0.5, 1.0, 1.5, and 2.0 kHz were broadcast using a CD player (Sony DE 356 K), connected to a 35-W amplifier (Alpine 3540) and a loudspeaker (Polk MM10a). Five broadcast tones were recorded with the microphones of two sound-level meters (Brüel & Kjaer 2238 and 2230) positioned at 0.5 and 8 m from the loudspeaker at ground level, respectively. Twelve propagation tests were conducted under different atmospheric conditions during day time.

At some sites, broadcast signals fell close to the background noise level at the farthest distance (8 m). In these cases, recorded signals were displayed at an expanded time scale, and the amplitude of the sinusoid was measured in short portions of the recording, to avoid measuring low-frequency fluctuations corresponding to background noise.

Air and substrate temperatures were measured after each recording with a thermocouple telethermometer (Fluke KJ 52) to the nearest 0.1 °C. Relative humidity was measured with a digital hygrometer (Oregon Scientific EM-913R).

D. Signal analysis

Amplitudes of the pure tones and playback calls recorded at different distances from the sound source were measured with SIGNALYZE 3.12 software in a Macintosh G4 computer.

E. Statistical analysis

Excess attenuation values of pure tones corresponded to single measurements at each site and distance. The excess attenuation values obtained for the six calls of a toad at a given distance and study site were converted to a linear scale (N/m^2) to calculate averages and then reconverted to dB. The average excess attenuation for calls of a species at a given distance and study site was also obtained following this procedure. The average excess attenuation values of calls and excess attenuation values of tones were log transformed to attain normality. Data were analyzed with STATISTICA 5 (StatSoft Inc.) software. Values of excess attenuation of pure tones and playback calls of individual toads across the distances and localities studied were considered as repeated values

For the propagation of pure tones, an ANOVA with three different factors was carried out. The factor "distance" considered the effect of the excess attenuation measured for tones of the same frequency for each of the four distances (1, 2, 4, and 8 m) as repeated values. The factor "locality", comprising ten levels, took into account the variation dependent on the different sites where the tests were conducted. Finally, the third factor, "frequency", was a continuous predictor of the model, considering spectral differences among the different pure tones used. This procedure was adopted to avoid excessive partitioning of the variance due to the high number ($N=27$) of pure-tone frequencies tested.

For the analysis of the differences in excess attenuation of the playback calls a different three-way ANOVA was carried out. The factors distance and locality were the same as in the model for pure tones, and the variation originated in the species-specific characteristics of the calls was assessed with the factor "species", including two levels (*A. cisternasii* and *A. obstetricans*). The values of dominant frequency of the calls of the different individuals of the two species were not included as a covariant in the model because more than 90% of variance of this spectral characteristic was explained by the species factor.

Since the main goal of this study was to explore adaptations of the calls of *Alytes* to the environment, differences in excess attenuation of the playback calls among localities according to the *Alytes* species native to each site were evaluated with planned comparisons. The comparison was therefore made between the localities for which *A. cisternasii* is native (Alcaçova, Grandola, Paço da Quinta, Laranjal, and Almadén de la Plata) and the localities for which *A. obstetricans* is native (Puente del Duque, Peñalara, Lago Enol, Covadonga, and Lago de la Cueva).

Significant differences among levels for factors having more than two levels were evaluated with Scheffé *post hoc* tests.

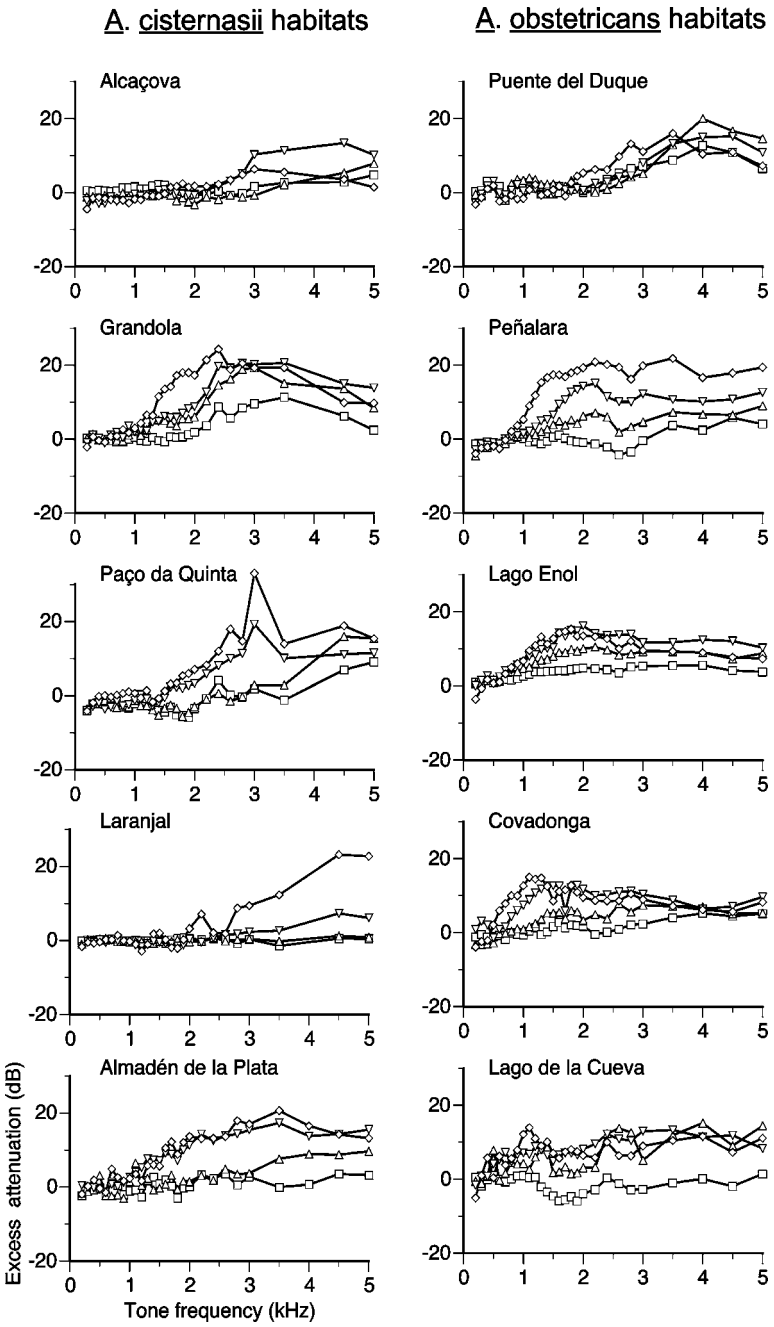


FIG. 2. Excess attenuation of pure tones measured in different habitats at 1 m (squares), 2 m (standing triangles), 4 m (inverted triangles), and 8 m (diamonds) relative to SPLs at 0.5 m from the loudspeaker.

III. RESULTS

A. Attenuation of pure tones

Measurements of propagation of pure tones showed that frequencies in the high range experienced higher rates of attenuation. The excess attenuation showed marked increases at frequencies starting at about 0.5 to 2 kHz, depending on the study site. High frequencies suffered the highest excess attenuation at 4 and 8 m, with measurements sometimes exceeding 15 dB (Fig. 2). Different patterns of excess attenuation were observed in sites to which *A. cisternasii* or *A. obstetricans* were native. For instance, an increase in excess attenuation was evident at about 2 kHz in Alcaçova and Laranjal (*A. cisternasii*) and Puente del Duque (*A. obstetricans*). By contrast, excess attenuation increased at about 1 kHz in Almadén de la Plata (*A. cisternasii*) and Peñalara, Lago Enol, Covadonga, and Lago de la Cueva (*A. obstetricans*). Differences between amplitudes of pure tones measured in nine localities at the beginning and at the end of a test showed small variation (average=0.08 dB, SD =2.43 dB, N=243 comparisons).

The ANOVA showed significant values for every factor considered (tone frequency, distance, and locality) and for the interactions among them (Table III). The frequency of pure tones had the greatest influence on the excess attenuation, higher frequencies experiencing the greatest excess attenuation. Excess attenuation of pure tones was also influenced by distance; however, significant differences occurred only between 1–4 m and 1–8 m (P=0.0066 and P =0.00066, respectively). Differences in excess attenuation of pure tones among localities were also highly significant.

TABLE III. ANOVA for excess attenuation of pure tones in the study sites of propagation of vocalizations of *A. cisternasii* and *A. obstetricans* in the Iberian Peninsula.

Factor	df	*F*	*P*
Frequency	1, 25	193.87	<0.0001
Distance	3, 75	3.84	0.0130
Locality	9, 225	18.10	<0.0001
Frequency/ Distance	3, 75	15.61	<0.0001
Frequency/ Locality	9, 225	6.65	<0.0001
Distance/ Locality	27, 675	9.92	<0.0001

However, for the entire range of pure-tone frequencies and distances, only the localities of Grandola and Lago Enol differed significantly from Alcacova, Paço da Quinta, and Laranjal ($P<0.0113$ and $P<0.0006$ in all cases for Grandola and Lago Enol, respectively). Covadonga differed significantly from Paço da Quinta ($P=0.0388$). Finally, all the interactions among factors in the model were highly significant. The relationship between excess attenuation and pure-tone frequency varied among distances and localities. Also, the relationship between excess attenuation and distances varied among localities (Table III).

To test for the influence of atmospheric conditions on sound propagation within the distance range used in the experiments in the Iberian Peninsula, 12 tone-propagation tests were conducted at a site in Chile during a single day from 7 AM to 5 PM at air temperature at ground level and humidity ranging from 8.4–26.4 °C and 50%–94%, respectively. During the day of the test, the atmosphere was calm, no detectable wind blew, and high clouds covered the sky. The excess attenuation showed restricted variation; the values measured for the 0.5, 1.0, 1.5, 2.0, and 3.0-kHz tones averaged 1.9 dB(range=1.3–2.4 dB), 10.2 dB(range=9.6–10.6 dB), 16.5 dB(range=15.7–17.8 dB), 23.2 dB(range=22.3–24.3 dB), and 22.0 dB(range=19.3–23.5 dB), respectively. The excess attenuation for the 0.5-kHz tone was significantly correlated with air temperature and humidity ($R^2=0.87$, $P<0.0001$ and $R^2=0.73$, $P<0.001$, respectively). The corresponding best-fit curves were $y=0.051\,x+0.93$ and $y=-0.019\,x+3.13$. The excess attenuation for the 1.5-kHz tone was marginally correlated with air temperature ($R^2=0.33$, $P=0.0507$) and the best fit curve was $y=-0.065\,x+11.77$. The rest of the relationships between excess attenuation of tones and air temperature and humidity were not statistically significant.

B. Attenuation of calls

Attenuation of playback calls followed patterns concordant with measurements of pure tones. Different patterns of attenuation were observed among localities native to both *A. cisternasii* and *A. obstetricans*. In three localities of *A. cisternasii* (Alcaçova, Paço da Quinta, and Laranjal), and one locality of *A. obstetricans* (Puente del Duque), calls of both species were subject to excess attenuation of less than 5 dB at all distances (Fig. 3). In contrast, in two localities of *A. cisternasii* (Grandola and Almadén de la Plata) and four localities of *A. obstetricans* (Peñalara, Lago Enol, Covadonga, and Lago de la Cueva) excess attenuation of about 10 dB was measured at 4 and 8 m for calls of *A. cisternasii* and *A. obstetricans*. In these sites, the attenuation of the calls of *A. cisternasii* was on average higher than those of *A. obstetricans*.

The peak SPLs measured for six repetitions of playback calls of an individual at a given distance during an experiment showed little variation, typically ±1 dB.

The ANOVA performed on the playback calls of the two species yielded high values of significance for every factor considered (species, distance, and localities) and for the interactions among them (Table IV). Distance was the factor having the highest influence on the excess attenuation of playback calls. The longer the distance tested, the higher the excess attenuation. Excess attenuation at every distance differed significantly from each other ($P=0.0481$ for the comparison between 4 and 8 m and $P<0.0001$ for all other comparisons). The excess attenuation also differed among species; the calls of *A. cisternasii* had the highest excess attenuation. Localities also differed in excess attenuation of playback calls, and most localities differed significantly with respect to excess attenuation of calls. Twelve out of 45 comparisons of excess attenuation between localities did not yield significant differences ($P>0.2265$ for all cases): Alcaçova versus Laranjal; Grandola versus Almadén de la Plata, Peñalara, and Lago de la Cueva; Almadén de la Plata versus Grandola, Lago Enol, Covadonga, and Lago de la Cueva; Puente del Duque versus Alcaçova and Peñalara; and Lago Enol versus Covadonga and Lago de la Cueva.

The interaction between distance and species was highly significant and the excess attenuation of calls of both *Alytes* species increased progressively with distance. In addition, the interactions between species and locality and between locality and distance were significant.

The planned comparisons performed with localities grouped according to the native species showed that the species' native condition of localities had a very important effect on call attenuation. The localities of *A. cisternasii* showed lower excess attenuation for the calls of both species altogether, relative to the localities of *A. obstetricans* ($F_{1,29}=2281.77$, $P<0.0001$; Fig. 4). *A. cisternasii* calls experienced significantly higher excess attenuation relative to the calls of *A. obstetricans* both in the localities of *A. cisternasii* ($F_{1,29}=20.89$, $P<0.0001$) and in the localities of *A. obstetricans* ($F_{1,29}=22.74$, $P<0.0001$). The interaction between species and native locality was also significant for the planned comparisons ($F_{1,20}=10.9$, $P=0.0026$). However, this interaction did not point to an optimal relationship between call transmission and native environment of either *Alytes* species: *A. cisternasii* calls experienced lower excess attenuation in the conspecific localities relative to the localities of *A. obstetricans* ($F_{1,29}=1434.43$, $P<0.0001$), but *A. obstetricans* calls experienced higher excess attenuations in the conspecific localities relative to the localities of *A. cisternasii* ($F_{1,29}=906.25$, $P<0.0001$).

Background noise measured during 15-s periods before the experiments of sound propagation yielded maxima rms

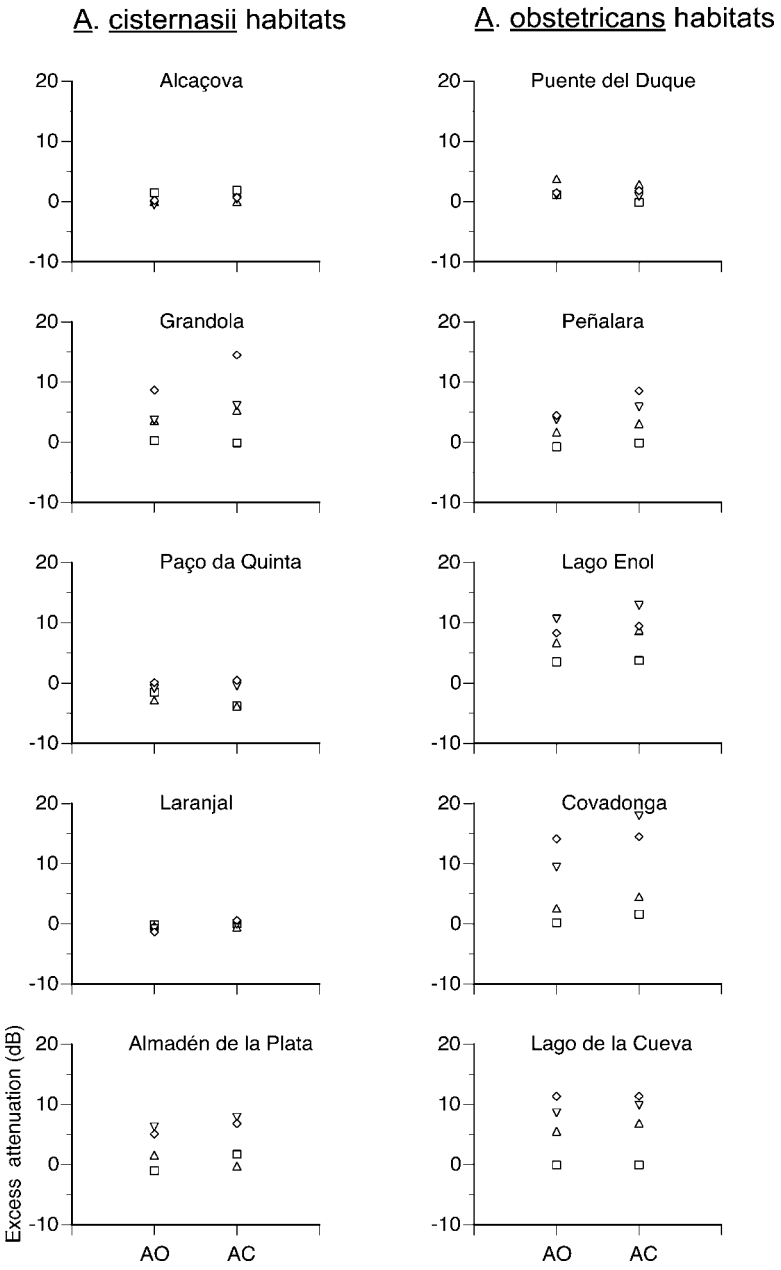


FIG. 3. Excess attenuation of playback calls of *Alytes cisternasii* and *A. obstetricans* measured in different habitats at 1 m (squares), 2 m (standing triangles), 4 m (inverted triangles), and 8 m (diamonds) relative to SPLs at 0.5 m from the loudspeaker. Each symbol represents the average for the calls of a species at a given distance. Abbreviations: AC: *A. cisternasii*, AO: *A. obstetricans*.

TABLE IV. ANOVA for excess attenuation of calls in the study sites of propagation of vocalizations of *A. cisternasii* and *A. obstetricans* in the Iberian Peninsula.

Factor	df	*F*	*P*
Species	1, 20	88.1	<0.0001
Distance	3, 60	1001.6	<0.0001
Locality	9, 180	428.4	<0.0001
Species/ Distance	3, 60	25.6	<0.0001
Species/ Locality	9, 180	9.9	<0.0001
Distance/ Locality	27, 540	163.4	<0.0001

SPL (fast weighting scale) ranging from 42–51 dB SPL in the different test sites. Environmental conditions during recordings are shown in Table I.

IV. DISCUSSION

The attenuation patterns of pure tones of frequencies above 1 kHz indicate that in most of the Iberian localities analyzed, the propagation of the calls of both species of *Alytes* will suffer a significant amount of excess attenuation. However, at some sites (Alcaçova, Laranjal, and Puente del Duque) we found that excess attenuation of tones below about 2.5 kHz was minimal, indicating that in these environments, propagation of tonal calls is not impaired. These different patterns are probably related to the proportions of various substrates in each site, having patches of sandy soil, rock, and herbaceous vegetation. Filtering properties of different substrates over which frogs communicate have been measured in South American (Penna and Solís, 1998) and

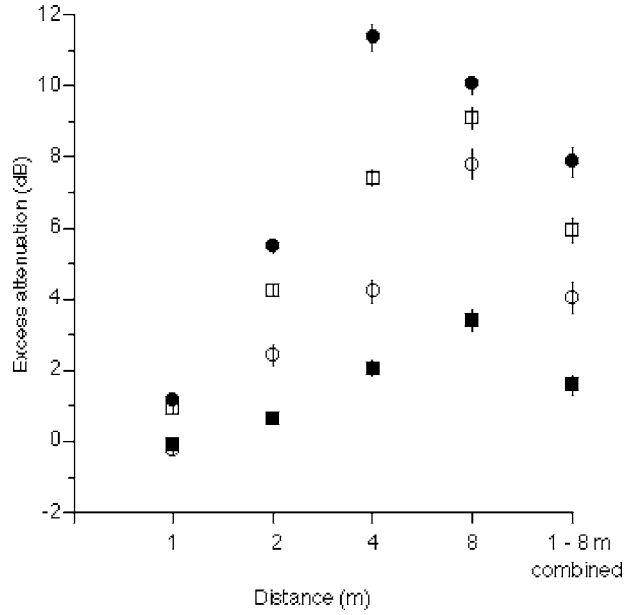


FIG. 4. Excess attenuation of the calls of *A. cisternasii* (circles) and *A. obstetricans* (squares) in conspecific (open symbols) and allospecific (filled symbols) localities. Symbols and bars indicate averages and standard errors for five localities, respectively.

North American (Venator, 1999) habitats. These studies have shown that attenuation of high frequencies occurs in open environments having absortive substrates such as bogs and grasslands.

The vocalizations of *Alytes* are soft sounds as compared to calls of other anurans, which in some cases exceed 100 dB rms SPL (e.g., Gerhardt, 1975; Loftus-Hills and Littlejohn, 1971; Passmore, 1981; Penna and Solís, 1998). Furthermore, the physical properties of the native environments of *Alytes*, producing excess attenuation of up to about 15 dB, constrain the range over which these toads communicate. This condition resembles those of species communicating across porous substrate in bogs in the austral temperate forest (Penna and Solis, 1998) and is at a disadvantage relative to anurans that communicate over water surface, where vocalizations do not experience or have negative excess attenuations (Penna and Solís, 1998; Boatright-Horowitz *et al.*, 1999).

The tone propagation tests conducted at a site in Chile indicate that the various atmospheric conditions under which the tests were conducted in different localities in the Iberian Peninsula are not likely to affect the propagation of the calls of *Alytes*. The restricted influence of temperature and humidity on excess attenuation, as indicated by the small ranges of variation and slopes of the best-fit curves (having maxima of 0.06 dB/°C and 0.01 dB/%, respectively), support this assumption.

Results of sound propagation experiments with broadcast calls show that the attenuation rates are related to the dominant frequency of the calls, *A. cisternasii* having higher attenuation rates relative to *A. obstetricans*. The calls of both species undergo attenuation rates that vary considerably, depending on the particular site tested. The analysis comparing habitats of both *Alytes* species shows that localities native to *A. cisternasii* are more favorable than habitats native to A. obstetricans for propagation of the calls of both species. This relationship is similar to the one reported for two subspecies of the North American cricket frog *Acris crepitans* native to open (*A. c. blanchardi*) and forest (*A. c. crepitans*) habitats. The calls of both taxa experience lower degradation (as measured with waveform autocorrelation) in the open relative to the forest habitat. The degradation of the calls of *A. c. crepitans* was lower in both open and forest habitats relative to *A. c. blanchardi* (Ryan *et al.*, 1990). These two studies illustrate that the relationship between the structure of a signal and the habitat in which it occurs is not always optimal. Rather, the species that occupies the less favorable habitat with respect to signal transmission produces signals that suffer less attenuation and/or degradation both in their native and nonnative habitats than the signals of the species occupying the more favorable habitat.

The distance over which midwife toads communicate can be estimated from the intensity of their vocalizations, their auditory thresholds, and the attenuation rates of their calls in natural environments. The calls of *A. cisternasii* and *A. obstetricans* are relatively soft sounds measuring about 78 dB rms SPL at 0.5 m from calling males (Marquez and Penna, unpublished data). The auditory thresholds of *A. cisternasii* measured with multiunit responses in the midbrain for frequencies in the range of the dominant frequencies of the conspecific advertisement calls are about 50 dB rms SPL for *A. cisternasii* (Bosch and Wilczynski, 2003; Bosch, Penna, and Márquez, unpublished data). The average excess attenuation of the calls of *A. cisternasii* in its own habitats at 1, 2, 4, and 8 m was −0.2. 2.4, 4.2, and 7.8 dB, respectively (see Fig. 4). These values generate a best-fit curve: $y=-11.54 \ln x+71.16$, in which x is distance from the loudspeaker (m) and y is the call intensity (dB rms SPL). This equation predicts that a call of 78 dB rms SPL at 0.5 m would reach 50 dB rms SPL at 6 m from the emitter, thus setting the maximum range over which males of this species would communicate and attract females.

The average excess attenuation of the calls of *A. obstetricans* in its own habitats was higher than the corresponding values for *A. cisternasii*: 1.1, 4.4, 7.6, and 9.4 dB at 1, 2, 4, and 8 m, respectively (see Fig. 4), which would restrict further the propagation range of the calls of this species. On the other hand, auditory thresholds of about 35 dB SPL have been reported for *A. obstetricans* (Mohneke and Schneider, 1979), a sensitivity that would facilitate signal detection in sound-attenuating environments relative to *A. cisternasii*. However, an accurate comparison of the auditory capabilities and communication ranges of both species requires electrophysiological recordings conducted in the same setup, which are currently underway.

Female *Alytes* give stronger phonotactical responses to calls having low dominant frequencies within the conspecific range (Márquez, 1995b; Márquez and Bosch, 1997a). However, there is evidence indicating a weaker frequency selectivity of female *A. cisternasii* as compared to *A. obstetricans* in complex experimental settings in which subjects have to choose between several synthetic calls of different frequencies (Márquez and Bosch, 1997a). It is likely that the transmission constraints derived from the relatively high frequency of the calls of *A. cisternasii* are not critical due to the

predominant explosive breeding modality of this toad, which limits the importance of phonotactical behavior for pairing.

In allopatric populations of these toads, females of both species prefer low-frequency calls (Márquez and Bosch, 1997b). The current study has shown that low-frequency calls suffer less attenuation in native habitats of either species. This may be an example of sexual and environmental selection pressures shaping together the evolution of the signal. A synergism between environmental and sexual selection pressures has been reported to occur for acoustic signals of African birds (Slabbekoorn and Smith, 2002). Where the toads occur in sympatry, female *A. cisternasii* prefer calls of average dominant frequency and this may be due to selection against mismatings with *A. obstetricans*. In these populations, the dominant frequencies of calls of the two species converge (Márquez and Bosch, 1997b), with *A. cisternasii* having lower dominant frequencies than when in allopatry and *A. obstetricans* having higher dominant frequencies. This is contrary to what would be expected if sexual selection alone were shaping dominant frequency of the calls in sympatric populations. The convergence of call dominant frequency is not consistent either with selection pressures for improved signal transmission in a common environment. Rather, this condition is congruent with evidence indicating that no strong environmental selective pressures have contributed to shaping the spectral design of anuran advertisement calls (Penna and Solís, 1998; Kime *et al.*, 2000). In the Iberian toads, interspecific spectral differences of the calls could result in part from selective pressures on body size. In particular, the relatively compact shape of *A. cisternasii* is likely to be an adaptation to fossorial habits in the xeric biotopes of southern Spain and Portugal (Crespo, 1982; Brown and Crespo, 2000).

The lack of a strict correspondence between the calls of *A. obstetricans* used for playbacks and those of populations of the test sites of Puente del Duque, Lago de la Cueva, and presumably Lago Enol and Covadonga, having higher dominant frequencies, indicate that these local populations may be subject to constraints for signal transmission that remain underestimated in our study.

Studies searching for relationships between call structure and habitat in anurans have shown no evidence for optimization of the kind occurring for birds. The lack of such matching in anurans is likely associated with the shorter range over which these vertebrates communicate relative to birds, their dispersion being restricted by water availability. Relationships between call structure and habitat may be unveiled by considering large number of taxa and a detailed characterization of their preferred habitats. In a comparative analysis of the calls of several species from Bolivia using such an approach, Bosch and de la Riva (2004) found that species calling in microhabitats characterized as closed environments produce calls containing more frequency modulations than species calling in open microhabitats. The exploration of the relationship between habitat and signal structure in taxa having calls of similar design but differing in their choice of habitat, as reported in the present study, contributes further insights into the environmental determinants of long-range signal structure. However, midwife toads are generalists in their habitat preferences and more defined relationships between environment and acoustic signals may be revealed by studies focused on species having habitat specializations.

ACKNOWLEDGMENTS

Travel of M. Penna to Portugal and Spain and of the Iberian authors to Chile was partially supported by Grants ICCTI/CONICYT 98-059, CSIC-CONICYT 2000-2-070, and by FONDECYT Grant 1010569. This study was also partially supported by travel grants between Lisbon and Madrid for R. Márquez and EG. Crespo [Consejo Superior de Investigaciones Científicas (CSIC) Convenio CSIC-ICCTI (Portugal) and Ministerio de Educación y Ciencia, Acción Integrada con Portugal, Proyecto HP2001-0076]. The Fonoteca Zoológica of the Museo Nacional de Ciencias Naturales (CSIC) in Madrid provided partial support for the sound analyses (Project 07M/0083/02 Comunidad de Madrid, P.I.R. Márquez). J. Bosch was supported by the Ramón y Cajal program from the Ministerio de Educación y Ciencia from Spain. This research complies with the regulations for native species protection and animal care of Spain and Portugal. Anonymous referees contributed comments that strengthened the final version of this manuscript.

Bee, M. A., and Perril, S. A. (**1996**). "Responses to conspecific advertisement calls in the green frog (*Rana clamitans*) and their role in male–male communication," Behaviour **133**, 283–301.

Boatright-Horowitz, S. S., Cheney, C. A., and Megela Simmons, A. (**1999**). "Atmospheric and underwater propagation of bullfrog vocalizations," Bioacoustics 9, 257–280.

Böll, S., and Linsenmair, K. E. (**1998**). "Size-dependent mating success and size assortative mating in the midwife toad *Alytes obstetricans*," Amphibia-Reptilia **19**, 75–90.

Bosch, J., and de la Riva, I. (**2004**). "Are frog calls modulated by the environment? An analysis with anuran species from Bolivia," Can. J. Zool. **82**, 880–888.

Bosch, J., and Márquez, R. (**1996**). "Acoustic competition in male midwife toads *Alytes obstetricans* and *Alytes cistenasii*: Response to neighbor size and calling rate. Implications for mate choice," Ethology **102**, 841–855.

Bosch, J., and Márquez, R. (**2001**). "Call timing in male-male acoustical interactions and female choice in the midwife toad *Alytes obstetricans*," Copeia **2001**, 169–177.

Bosch, J., and Wilczynski, W. (**2003**). "Auditory tuning of the Iberian midwife toad, *Alytes cisternasii*," Herp. J. **13**, 53–57.

Brown, L. E., and Crespo, E. G. (**2000**). "Burrowing behavior of the midwife toads *Alytes cisternasii*, and *A. obstetricans* (Anura, Discoglossidae)," Alytes (ISSCA) **17**, 101–113.

Crespo, E. G. (**1982**). "Contribuição para o conhecimento da biología das espécies Ibéricas de *Alytes, Alytes obstetricans* Boscai (Lataste 1879) e *Alytes cisternasii* (Boscá 1879) (Amphibia Discoglossidae). Ovos, posturas (Epocas de reprodução)," Arquivos do Museo Bocage. Serie A **1**, 453–466.

Crespo, E. G., Oliveira, M. E., Rosa, H. C., and Paillette, M. (**1989**). "Mating calls of the Iberian midwife toads *Alytes obstetricans Boscai* and *Alytes cisternasii*," Bioacoustics 2, 1–9.

Gerhardt, H. C. (**1975**). "Sound pressure levels and radiation patterns of the vocalizations of some North American frogs and toads," J. Comp. Physiol. [A] **102**, 1–12.

Gerhardt, H. C., and Huber, F. (**2002**). *Acoustic Communication in Insects and Anurans* (The University of Chicago Press, Chicago).

Gish, S. L., and Morton, E. S. (**1981**). "Structural adaptations to local habitat acoustics in Carolina wren songs," Z. Tierpsychol. **56**, 74–84.

Harris, C. M. (**1966**). "Absorption of sound in air versus humidity and temperature," J. Acoust. Soc. Am. **40**, 148–159.

Kime, N. M., Turner, W. R., and Ryan, M. J. (**2000**). "The transmission of advertisement calls in Central American frogs," Behav. Ecol. **11**, 71–83.

Loftus-Hills, J. J., and Littlejohn, M. J. (**1971**). "Mating call sound intensities of anuran amphibians," J. Acoust. Soc. Am. **49**, 1327–1329.
López, P. T., Narins, P. M., Lewis, E. R., and Moore, S. W. (**1988**). "Acousticallly induced modification in the white-lipped frog *Leptodactylus albilabris*," Anim. Behav. **36**, 1295–1308.
Márquez, R. (**1992**). "Terrestrial paternal care and short breeding seasons: Reproductive phenology of the midwife toads *Alytes obstetricans*, and *A. cisternasii*," Ecography **15**, 279–288.
Márquez, R. (**1995a**). "Female choice in the midwife toads (*Alytes obstetricans* and *A. cisternasii*)," Behaviour **132**, 151–161.
Márquez, R. (**1995b**). "Preferencia de las hembras por cantos de frecuencia dominante baja en el sapo partero común *Alytes obstetricans* (Anura, Discoglossidae). Experimentos *in situ*," Rev. Española Herpetol. **9**, 77–83.
Márquez, R., and Bosch, J. (**1995**). "Advertisement calls of the midwife toads *Alytes* (Amphibia, Anura, Discoglossidae) in continental Spain," J. Zool. Syst. Evol. Res. **33**, 185–192.
Márquez, R., and Bosch, J. (**1997a**). "Female preference in complex acoustical environments in the midwife toads *Alytes obstetricans* and *Alytes cisternasii*," Behav. Ecol. **8**, 588–594.
Márquez, R., and Bosch, J. (**1997b**). "Male advertisement call and female preference in sympatric and allopatric midwife toads (*Alytes obstetricans* and *Alytes cisternasii*)," Anim. Behav. **54**, 1333–1345.
Márquez, R., and Crespo, E. G. (**2002**). "*Alytes cisternasii*," in *Atlas y Libro Rojo de los Anfibios y Reptiles de España*, edited by J. M. Pleguezuelos, R. Márquez, and M. Lizana (Dirección General de Conservación de la Naturaleza. Ministerio de Medio Ambiente-Asociación Herpetológica Española, Madrid), pp. 73–75.
Mohneke, R., and Schneider, H. (**1979**). "Effect of temperature upon auditory thresholds in two anuran species, *Bombina v. variegata* and *Alytes o. obstetricans* (Amphibia: Discoglossidae)," J. Comp. Physiol. [A] **130**, 9–16.
Morton, E. S. (**1975**). "Ecological sources of selection on avian sounds," Am. Nat. **109**, 17–34.
Passmore, N. I. (**1981**). "Sound levels of mating calls of some African frogs," Herpetologica **37**, 166–171.
Penna, M., and Solís, R. (**1998**). "Frog call intensities and sound propagation in the South American temperate forest region," Behav. Ecol. Sociobiol. **42**, 371–381.
Piercy, J. E., and Daigle, G. A. (**1991**). "Sound propagation in the open air," in *Handbook of Acoustical Measurements and Noise Control*, edited by C. M. Harris (McGraw-Hill, New York), pp. 3.1–3.26.
Reading, C. J., and Clarke, R. T. (**1988**). "Multiple clutches, egg mortality and mate choice in the mid-wife toad, *Alytes obstetricans*," Amphibia-Reptilia **9**, 357–364.
Rodríguez-Jiménez, A. J. (**1984**). "Fenología del sapo partero ibérico (*Alytes cisternasii* Boscá. 1879)," Alytes (España) **2**, 9–23.
Ryan, M. J., and Brenowitz, E. A. (**1985**). "The role of body size, phylogeny and ambient noise in the evolution of bird song," Am. Nat. **126**, 87–100.
Ryan, M. J., Cocroft, R. B., and Wilczynski, W. (**1990**). "The role of environmental selection in intraspecific divergence of mate recognition signals in the cricket frog, *Acris crepitans*," Evolution (Lawrence, Kans.) **44**, 1869–1872.
Shy, E. (**1983**). "The relation of geographical variation in song to habitat characteristics and body size in North American tanagers (*Thraupinae: Piranga*)," Behav. Ecol. Sociobiol. **12**, 71–76.
Slabbekoorn, H., and Smith, T. B. (**2002**). "Habitat-dependent song divergence in the little greenbull: An analysis of environmental selection pressures on acoustic signals," Evolution (Lawrence, Kans.) **36**, 1849–1858.
Sorjonen, J. (**1986**). "Song structure and singing strategies in the genus *Luscinia* in different habitats and geographical areas," Behaviour **98**, 274–285.
Van Staaden, M. J., and Römer, H. (**1997**). "Sexual signalling in bladder grasshoppers: Tactical design for maximizing call range," J. Exp. Biol. **200**, 2597–2608.
Venator, K. (1999). "The influence of signal attenuation and degradation on behavior and midbrain auditory thresholds in the cricket frog, *Acris crepitans* Blanchardi," Unpublished doctoral dissertation, University of Texas at Austin.
Wagner, W. E., Jr. (**1989**). "Fighting, assessment, and frequency alteration in Blanchard's cricket frog," Behav. Ecol. Sociobiol. **25**, 429–436.
Zimmerman, B. L. (**1983**). "A comparison of structural features of calls of open and forest habitat in the central Amazon," Herpetologica **39**, 235–246.

Noninvasive generation and measurement of propagating waves in arterial walls

Xiaoming Zhang[a] and James F. Greenleaf
Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, 200 First Street SW, Rochester, Minnesota 55905

(Received 1 August 2005; revised 21 November 2005; accepted 27 November 2005)

Arterial wall stiffness can be associated with various diseases. Arteriosclerosis involves the buildup of plaques within artery walls that stiffen the arteries. The stiffness of an artery can be assessed by measurement of the pulse wave velocity (PWV). Usually, PWV is estimated using the foot-to-foot method. However, the foot of the pressure wave is not very clear due to reflected waves. Also, the blood pressure wave generated by the heart is normally a low frequency wave, hence the time resolution is low. PWV is an average indicator of artery stiffness between the two measuring positions, and therefore cannot easily identify local stiffness. In this paper a force on the arterial wall is generated noninvasively by the radiation force of ultrasound. Techniques for measuring the propagating wave due to this localized force are studied. The excitation force can be either a very short pulse or a modulated cw signal of a few hundred hertz. The temporal resolution of this method, which is in the range of microseconds, is much higher than the conventional pressure PWV method, and therefore allows the wave velocity to be measured accurately over short distances of a few millimeters. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159294]

PACS number(s): 43.80.Jz, 43.80.Qf, 43.80.Vj, 43.20.Hq [CCC] Pages: 1238–1243

I. INTRODUCTION

According to recent statistics by the American Heart Association, cardiovascular disease (CVD) has been the number one killer in the United States.[1] CVD claims more lives each year than the next five leading causes of death combined, which are cancer, chronic lower respiratory diseases, accidents, diabetes mellitus, and influenza and pneumonia. It has long been recognized that a high percentage of all cardiovascular disease is associated with a stiffening of the arteries or arteriosclerosis.[2] Arteriosclerosis involves the buildup of plaques on the insides of the artery walls, which causes thickening and hardening of the arteries. It can ultimately lead to high blood pressure and to angina (chest pain), heart attack, stroke, and/or sudden cardiac death. It is well known that a stiffened artery can be associated with various diseases and with aging.[2–4] Therefore increased stiffness of the arteries has recently gained acceptance as a potential risk factor for cardiovascular and many other diseases.[5–8]

Pulse wave velocity (PWV) is widely used for estimating the stiffness of an artery.[9–11] Measurements of pulse wave velocity in man, starting with the classic work of Hallock[2] (on 500 persons, aged 5–85), have shown that the velocity increases with age, from an average of 4 m/s in young persons to 10 m/s in old persons. (The values are for the aorta from heart to subclavian artery.) Measurement of the PWV using ultrasound has been an area of recent interest.[12–14]

The velocity of propagation of the arterial pulse is dependent on the dimensions and stiffness (modulus of elasticity) of the intervening vessels. PWV is defined as the distance between two measurement points divided by the pulse wave transit time-delay from the two points[15]

$$c_p = L/T, \tag{1}$$

where L is the length of two measuring points and T is the transit time of the pulse between these two points. The pulse wave velocity is directly related to the Young's modulus in the circumferential direction of the artery by the well-known Moens-Korteweg equation (1878).[3] This equation was modified by Bramwell[16] in which the pulse wave velocity is related to a single parameter "arterial distensibility" which can be measured conveniently in man.

Despite the simple definition of PWV, some problems still remain that limit the interpretation of the available data and the general applicability of PWV measurement.[17,18] Usually, PWV is measured using the "foot-to-foot" method.[13] The "foot" of the pressure wave is not clear due to reflected waves and blood noise. In addition, the pressure pulse, produced by the pumping of the heart, is predominately a low frequency wave. Therefore, it is difficult to measure small time changes from the pulse wave forms. In other words, the time resolution of the pulse is low, therefore, a relatively large distance is needed for reliable time change measurement. Measurement over a large distance means PWV produces an average value of artery stiffness between the two measuring points. It is therefore difficult to identify local stiffness variation of a few millimeters length. For early diagnosis of arteriosclerosis, a measurement of local stiffness in a few millimeters length is needed[19] because in the early stage of arteriosclerosis, a fibrous spot is several millimeters in diameter and becomes homogeneously hard in the final stage.

In this paper an external force is generated noninvasively in the arterial wall by localized radiation force of ul-

[a]Author to whom correspondence should be addressed; electronic mail: zhang.xiaoming@mayo.edu

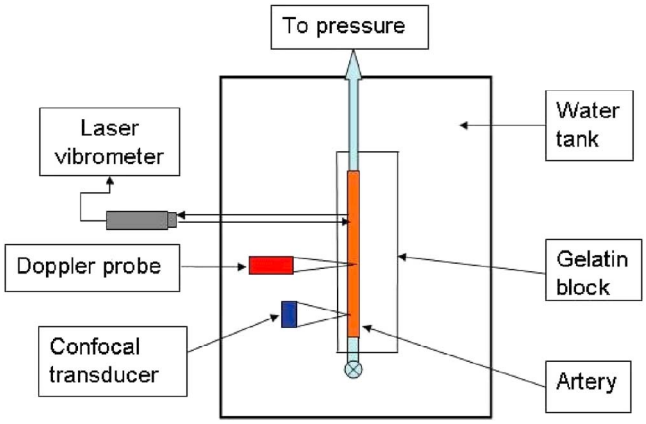


FIG. 1. (Color online) Schema of the experimental system for measuring the wave velocity in an artery.

trasound. Techniques for measuring the propagation velocity of the resulting wall waves are studied for either a short pulse or a modulated cw wave. The present method offers a new tool for measuring the wave velocity of an artery over a short length, which can help to evaluate "local" elastic modulus of artery.

II. METHOD

Our method is to use the radiation force of ultrasound to generate either a short pulse or a modulated cw wave of a few hundred hertz in an artery and measure the wave velocity of its propagation. The duration of the pulse is very short, for example, 0.1 ms if 300 cycles of 3 MHz toneburst is used, therefore, the time resolution is quite high, allowing the time shift of the pulse wave to be measured very accurately in the microsecond range. A low frequency cw wave in a few hundred hertz is generated by interfering two ultrasound beams of slightly different frequencies. The wave velocity in the artery can then be measured at the difference frequency of the two ultrasound beams that is of the order of a few hundred hertz.

Several methods are studied for measuring the wave velocity in the artery. Among those a scanning technique is studied in which the detector is fixed and the source is scanned. With the scanning technique two- and three-dimensional images of wave propagation can be obtained from which the wave velocity is calculated.

The experiments were conducted in a water tank. An artery is embedded in a gelatin phantom. The artery can be pressurized through the saline inside the artery. The radiation force of ultrasound is used to generate a wave propagating along the artery. The wave propagation is measured either with laser or a Doppler transceiver. A schema of this experimental system is shown in Fig. 1.

A. Phantom preparation

The measurement techniques are studied first in a rubber tube and then applied to an artery. To prepare a tube phantom, a 24 cm length of 5 mm outer radius by 3 mm inside radius latex rubber tubing (Kent Elastomer Products, Inc., OH) was secured to rigid connectors at each end and mounted in a three-sided aluminum/acrylic frame. The tube was slightly stretched lengthwise to place it under tension and prevent sagging. The tube axis was positioned at the center of the frame's 6×6 cm cross section. The two long open sides of the frame were temporarily sealed and a gelatin mixture was poured into the frame, filling it completely. The gelatin mixture had been prepared by dissolving 60 g of porcine gelatin (Sigma, G2500, 300 bloom) in a 60/40 solution of water and glycerol to make a 1000 ml volume, with 1% potassium sorbate added as a preservative. The phantom was then allowed to rest overnight at room temperature. An artery phantom was prepared in the same way in which an 8 cm portion of fresh porcine femoral artery was secured with sutures to a length of 2 mm outer diameter silicone tubing at each end. The artery or tubing were anchored inside an aluminum frame and pressurized with normal saline to 60 mm Hg. An open side of the frame was sealed with a thin plastic membrane and a gelatin mixture was poured into the frame, filling it completely. The artery was positioned near the open side of the frame such that it lay approximately 2 mm below the final surface of the gelatin. The gelatin mixture had been prepared as the above for the rubber tube. The artery remained pressurized until the gelatin had set up, after which the artery phantom was placed in a refrigerator overnight.

B. Wave generation

The radiation force of ultrasound is used to generate a propagating wave in the tube or the artery. Two kinds of signals are studied in this paper. One is to generate a very short pulse, and another is a modulated cw wave in a few hundred hertz. The very short pulse is directly generated by the ultrasound beam. For enough force amplitude, a 3 MHz tone burst of a few hundred cycles is used. For generating a low frequency signal two intersecting cw focused ultrasound beams of slightly different frequencies are used to generate a localized oscillating radiation force.[20–22] This force is modulated at the difference frequency of the two ultrasound beams, thus, a low frequency localized force is produced which excites the arterial wall.

C. Wave velocity measurement

Wave propagation in the tube or artery was measured noncontact by a laser vibrometer with a Polytec VibraScan Laser Vibrometer[23] system. This laser vibrometer uses the helium neon laser of 633 nm wavelength and the Doppler effects to measure the vibration of an object. The vibrometer has large dynamic sensitivity. It can measure the vibration frequency from 0.3 Hz to 250 kHz and the vibration velocity from 0.3 μm/s to 10 m/s. The laser scanner can scan at prescribed two-dimensional grids. The maximum deflection of the scanner is ±20° in both the horizontal and vertical directions. The angular resolution is less than 0.002°, which results in a 9 μm spatial resolution for an object at 0.5 m away.

A scanning technique was studied for measuring the wave velocity of the artery in which the laser vibrometer is fixed and the ultrasound force transducer is scanned. Although the measurement of excised or superficial arteries can

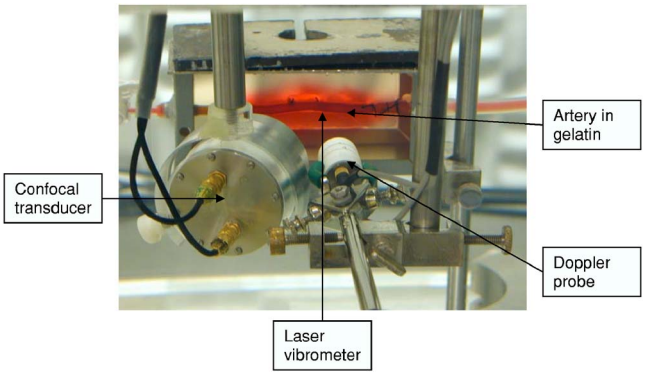


FIG. 2. (Color online) Typical experimental setup in which a confocal ultrasound transducer is at the left. An excised artery is embedded in a tissue-mimicking gelatin phantom. The artery can be pressurized by internal saline through a silicone tube. The gelatin block is in a water tank. The transducer remotely produces a localized force at the artery. The wave propagation in the artery is measured noncontact by a laser vibrometer or a Doppler probe.

be done with laser, the measurement of deep arteries or *in vivo* measurements is done with ultrasound Doppler techniques. A cw Doppler transceiver was also used to measure the wave velocity of rubber tubes and arteries in gelatin. These techniques are suitable for clinic applications. Recently, we have successfully applied these techniques for *in vivo* noninvasively measuring the wave velocity of femoral arteries of pig.[24] A typical *in vitro* experiment setup is shown in Fig. 2.

III. RESULTS AND DISCUSSION

Experiments were carried out on a silicone rubber tube in gelatin and an artery in gelatin. The wave velocity of a short pulse was first measured by a convention technique, "fixed source and scanning detector," in which the ultrasound force transducer is fixed and the laser detector is moved to measure the pulse waves at different positions. Then a new technique, "scanning source and fixed detector," was studied in which the ultrasound force transducer is scanned and the laser detector is fixed to measure the wave propagation for the tube and the artery. In the second technique, a custom-built commercial three-axis scanning system (Brandt Automation, Delano, MN) is used. The scanner can perform linear scans in each of the three axes with a resolution of 2.5 μm and at speeds over 25 cm/s. Scanning motion is controlled by an intelligent servo controller (Galil Motion Control, Rocklin, CA) via in-house developed LABVIEW VI's (National Instruments, Austin, TX). Experimental signals are digitized by a 14-bit, 100 MS/s two-channel digitizer (Alazartech, Montreal, QC, Canada) and displayed on a personal computer.

A. Fixed source and scanning detector

Figure 3 shows two pulse response curves of the tube to a 3 MHz toneburst of 300 cycles at two positions close to the force source. If we measure the propagation of the induced pulse at a known distance away from the source in the longitudinal direction of the tube, we can calculate the velocity of the pulse in the longitudinal direction of the tube. One curve shows the pulse response at the first position, while

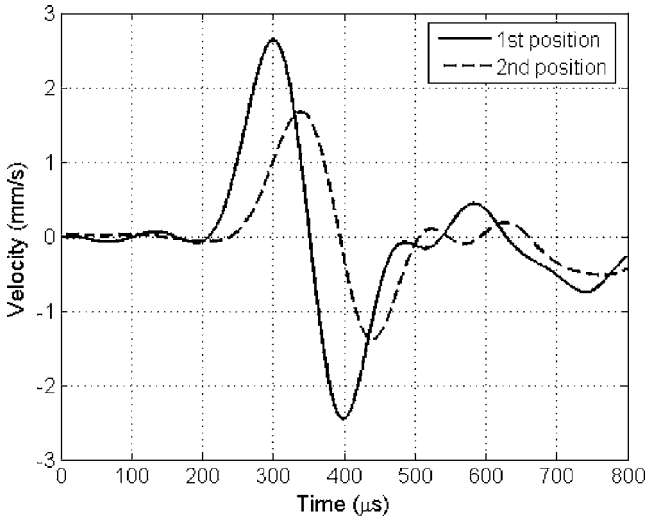


FIG. 3. Two pulse wave responses in a silicone rubber embedded in gelatin to a short pulse force generated by ultrasound at two positions with 2 mm distance.

another curve shows the pulse response at the second position 2 mm away from the first measurement position. We can use a characteristic point in the pulse shapes to measure the transit time, for example, the first positive peak or the first negative peak as a reference point. The first peak appears at 299 and 339 μs, respectively, in Fig. 3. The transit time can be measured between the difference of the time from the peak point of the pulse shape and is 40 μs for the pulse to travel 2 mm, therefore, the pulse velocity is calculated as 2 mm/40 μs=50 m/s.

It is important to mention that we can calculate the same value of pulse velocity if we measure transit time using the negative peak of the pulse shape. This means that the pulse keeps its shape when traveling a short distance of 2 mm. Normally, the arterial pulse gradually changes its shape when it propagates. This is because of reflected waves due to material and geometric changes in the artery. In the classically used "foot-to-foot" technique, the pulse wave velocity is measured by the time of travel of the "foot" of the pressure wave over a known distance. The "foot" is defined as the point, at the end of diastole, when the steep rise of the wave front begins. By analogy with the interaction of a transient wave with reflected waves that it creates, the early part of the wave will be little affected by the reflections.[3] Therefore, it is reasonable to treat the early wave front as a region that maintains its identity in the propagated wave. The "foot" of the wave front is, at first sight, the easiest point to recognize; it is, however, clear-cut only when the wave is recorded at a relatively long distance. It becomes much more difficult to define with any precision for recording over short distances.

Pulse shape change also comes from nonlinearity of material property of artery and surrounding tissues. This is more obvious when the pulse travels a long distance. The change of pulse shape makes the characteristic point less reliable, which causes significant errors in measurement of pulse transit time. The widely used "cross-correlation" method is also affected by the pulse shape change between the two measurement points.[17] The advantage of our method is that we

can measure the pulse velocity over distance as short as 2 mm. We can therefore avoid the effects of reflected waves and material nonlinearity.

Another advantage is that we use a very short pulse compared to the blood pressure pulse produced by the heart. For example, the pulse duration for the experiment is 100 μs. We can easily identify a 1 μs time shift, therefore the time resolution is in the range of μs. The blood arterial pulse duration is close to 1 s, and the ability to differentiate time changes is typically in the range of milliseconds. Therefore, the time resolution of our method is several orders higher than that of blood arterial pulse. This is not surprising because we use a toneburst of megahertz signals, but the blood arterial pulse, produced by the heart, is predominately a low frequency wave of a few hertz.

B. Scanning source and fixed detector

Figure 4(a) is a three-dimensional graph of pulse wave propagation in which the three coordinates are, respectively, time, distance, and amplitude. The time resolution is 10 μs, and the total sampled time is 1.28 ms. The distance resolution is 0.1 mm, and the total scanned length is 12.8 mm. The signal amplifier gain was adjusted so that the wave amplitude signals were digitized between ±2048. To obtain this figure, we fixed the laser detector and scanned the source transducer 128 points at 0.1 mm intervals. For each point, we digitized the laser velocity signal at 100k samples/s.

The velocity of the pulse can be calculated from any two pulse shapes at two different positions. Any horizontal line in Fig. 4(a) represents a pulse wave at a specific position. Although in Fig. 4(a) we have even finer distance resolution, 0.1 mm, we do not necessarily need such fine resolution. In the "scanning source and fixed detector" technique, a much finer resolution can be obtained than that of the "fixed source and scanning detector." This is because the force transducer can be scanned at a resolution of 2.5 μm, while a ruler is used for measuring the distance when scanning the detector transducer. If we still use 2 mm distance resolution as before and take the first and 21st horizontal lines in Fig. 4(a), we can draw a graph as shown in Fig. 4(b). The transit time between the two curves in Fig. 4(b) can be measured by the time shift of either the first positive or first negative peaks of the wave shape which is measured to be 40 μs. Therefore, the pulse velocity is 2 mm/40 μs=50 m/s, which is the same as the "fixed source and scanning detector" technique.

The three-dimensional graph of Fig. 4(a) can be also shown in a two-dimensional image as of Fig. 4(c) in which the intensity of the image is the wave amplitude. The slope of the peak or the valley of pulse wave in the image is actually the wave velocity of the pulse. The steeper the slope, the higher the velocity.

A cw Doppler transceiver was used to measure the wave velocity of the tube for different low frequencies. We fixed the Doppler transceiver and scanned the ultrasound radiation force transducer to measure the wave velocity. Figure 5 shows two images of the results for the tube in gelatin for frequencies of 250 and 450 Hz, respectively. The time coordinate of the image is with resolution of 20 μs. The distance

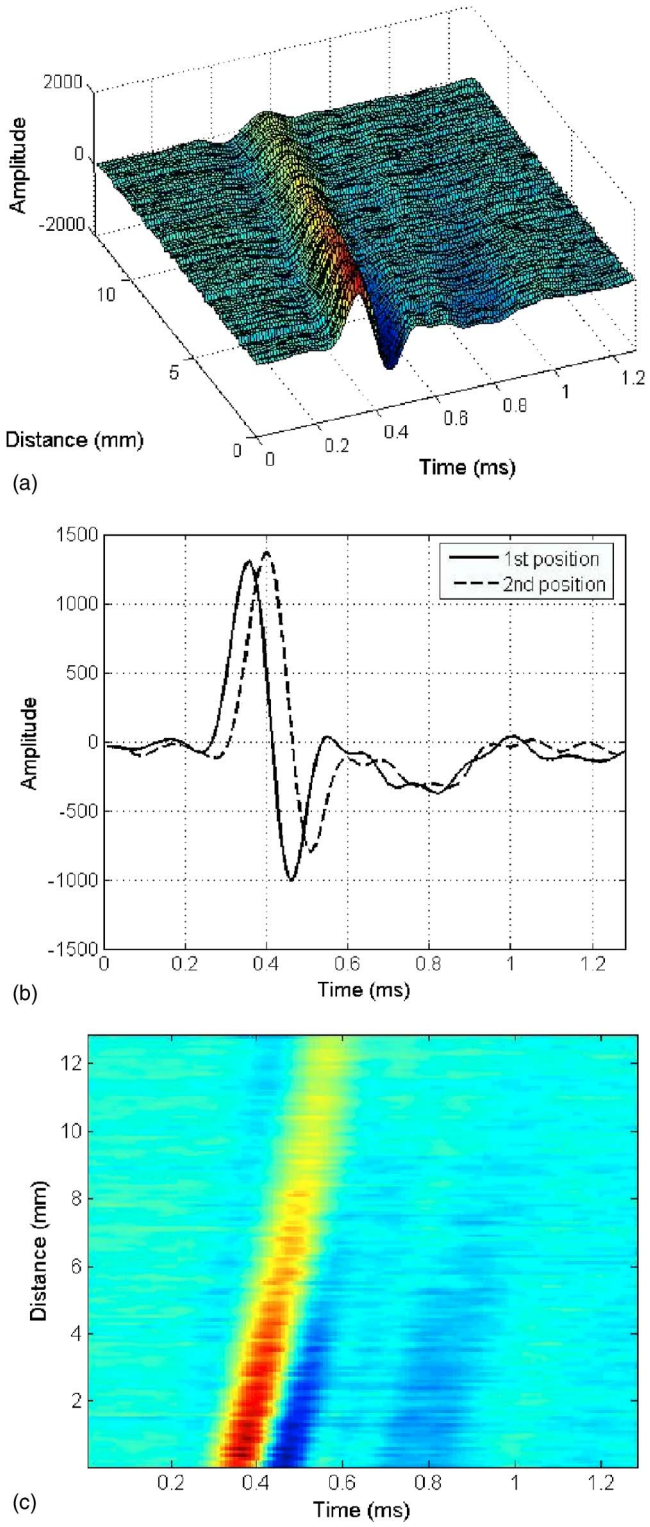


FIG. 4. (Color online) Graphs of pulse wave propagation in the rubber tube embedded in gelatin. The time resolution is 10 μs, the distance resolution is 0.1 mm, and the wave amplitude signals are digitized between ±2048; (a) three-dimensional graph; (b) two pulse shapes at two positions with 2 mm distance; (c) two-dimensional image.

coordinate of the image is with resolution of 0.1 mm. From the slope of each image the wave velocity can be measured. The wave velocity is measured 13.5 m/s for 250 Hz and 17.8 m/s for 450 Hz. By measuring the wave velocity at different frequencies, the dispersion of wave velocity with

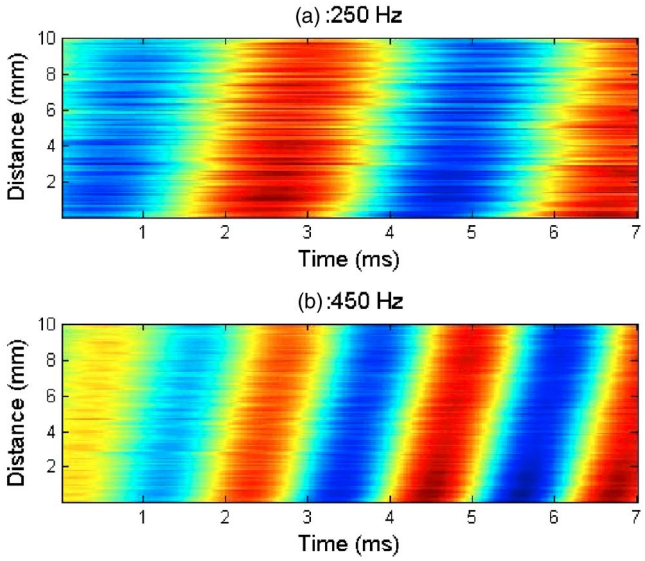


FIG. 5. (Color online) Two images of the wave propagation of the silicone tube in gelatin for frequencies of 250 and 450 Hz, respectively. The time resolution of the figure is 20 μs. The distance resolution of the figure is 0.1 mm. From the slope of the image the wave velocity can be measured. The wave velocity is measured 13.5 m/s for 250 Hz and 17.8 m/s for 450 Hz.

frequency can be evaluated. The artery wave dispersion and its change might be used to identify artery diseases and disease progression. More results will be needed for wave dispersion of diseased arteries to find the sensitivity of wave dispersion on artery disease. The Doppler technique is potentially suitable for clinical applications.

We have shown the two scanning techniques for measuring the wave velocity in a tube under the excitation of ultrasound. One is to fix the source and scan the detector, the other one is to scan the source and fix the detector. Although the two techniques are different in scanning method, they are actually the same in physics and mathematics if we use a different coordinate as the reference coordinate. Both techniques provide the same results. The first technique is simpler, but the second can provide finer distance resolution and images of the pulse wave propagation.

C. Experiments on an artery

The scanning technique was applied for measuring the wave velocity of short pulse in an artery. Figure 6 is a three-dimensional graph of pulse wave propagation in the artery embedded in a gelatin block. A 900 μs toneburst of 3 MHz was used to excite the artery. The time resolution of the figure is 10 μs, and the total sampled time is 2.56 ms. The distance resolution is 0.05 mm and the total scanned length is 5 mm. The wave decays very quickly in the artery. The time shift is measured 0.5 ms from the first peak for a propagating length of 5 mm, therefore the wave velocity of the pulse calculated from this graph is 10 m/s for the artery embedded in gelatin.

IV. CONCLUSION

In this paper an external force is generated noninvasively on the arterial wall by the radiation force of ultrasound. A new scanning technique, termed "scanning source, fixed detector," is studied for measuring the wave velocity in arteries under the excitation of ultrasound. This technique can be used for measuring the wave velocity over a short length of a few millimeters, which helps to assess the local stiffness or elastic modulus of the artery. This technique provides a more accurate means for measuring either a short pulse wave velocity or a modulated low frequency cw wave velocity in arteries, which avoids the effects of wave reflection and nonlinearity. With the Doppler detection, this technique can be used for measuring local artery stiffness, which may be useful in identifying artery diseases at an early stage, characterizing the disease and its progress. This technique has been recently applied for *in vivo* noninvasive measurement of local wave velocity in femoral arteries of pig.[24]

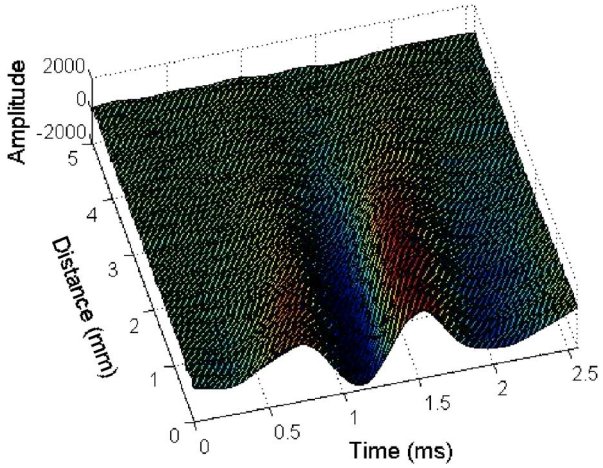


FIG. 6. (Color online) A three-dimensional graph of pulse wave propagation in the artery embedded in gelatin. The time resolution is 10 μs, the distance resolution is 0.05 mm, and the wave amplitude signals are digitized between ±2048. The wave velocity is measured 10 m/s for the artery embedded in gelatin.

ACKNOWLEDGMENTS

This study is supported by Grant No. EB 02640 from the National Institutes of Health. The authors thank R. R. Kinnick for experimental assistance.

[1]American Heart Association, *Heart Disease and Stroke Statistics—2003 Update*, (American Heart Association, Dallas, 2003).
[2]P. Hallock, "Arterial elasticity in man in relation to age as evaluated by the pulse wave velocity method," Arch. Intern Med. **54**, 770–798 (1934).
[3]W. W. Nichols and M. F. O'Rourke, *McDonald's Blood Flow in Arteries*, 3rd ed. (Arnold, London, 1990).
[4]K. Sutton-Tyrrell, R. H. Mackey, R. Holubkov, P. V. Vaitkevicius, H. A. Spurgeon, and E. G. Lakatta, "Measurement variation of aortic pulse wave velocity in the elderly," Am. J. Hypertens. **14**, 463–468 (2001).
[5]M. F. O'Rourke, J. A. Staessen, C. Vlachopoulos, D. Duprez, and G. E. Plante, "Clinical application of arterial stiffness; definitions and reference values," Am. J. Hypertens. **15**, 426–444 (2002).
[6]L. M. V. Bortel, D. Duprez, M. J. Starmans-Kool, M. E. Safer, C. Giannattasio, J. Cockcroft, D. R. Kraiser, and C. Thuillez, "Clinical application of arterial stiffness; task force III: Recommendations for user procedures," Am. J. Hypertens. **15**, 445–452 (2002).
[7]G. M. London and J. N. Cohn, "Prognostic application of arterial stiffness: Task forces," Am. J. Hypertens. **15**, 754–758 (2002).
[8]S. Laurent, S. Katsahian, C. Fassot, A. Tropeano, I. Gautier, B. Laloux, and P. Boutouyrie, "Aortic stiffness is an independent predictor of fatal stroke in essential hypertension," Stroke **34**, 1203–1206 (2003).

[9]A. Steptoe, H. Smulyan, and B. Cribbin, "Pulse wave velocity and blood pressure change: Calibration and applications," Psychophysiology **13**, 488–493 (1976).
[10]J. D. Pruett, J. D. Bourland, and L. A. Geddes, "Measurement of pulse-wave velocity using a beat-sampling technique," Annals Biomech. Eng. **16**, 341–347 (1988).
[11]Y. Nagai, J. L. Fleg, M. K. Kemper, T. M. Rywik, C. J. Earley, and E. J. Metter, "Carotid arterial stiffness as a surrogate for aortic stiffness: Relationship between carotid artery pressure-strain elastic modulus and aortic pulse wave velocity," Ultrasound Med. Biol. **25**, 181–188 (1999).
[12]P. J. Brands, J. M. Willigers, L. A. F. Ledoux, R. S. Reneman, and A. P. G. Hoeks, "A noninvasive method to estimate pulse wave velocity in arteries locally by means of ultrasound," Ultrasound Med. Biol. **24**, 1325–1335 (1998).
[13]J. M. Meinders, L. Kornet, P. J. Brands, and A. P. G. Hoeks, "Assessment of local pulse wave velocity in arteries using 2D distension waveforms," Ultrason. Imaging **23**, 199–215 (2001).
[14]A. Eriksson, E. Greiff, T. Loupas, M. Persson, and P. Preque, "Arterial pulse wave velocity with tissue Doppler imaging," Ultrasound Med. Biol. **28**, 571–580 (2002).
[15]A. L. King, "Circulatory system: Arterial pulse; wave velocity," in *Medical Physics*, editor-in-chief, O. Glasser (Year Book Publishers, Chicago, 1950), Vol. **II**.
[16]J. C. Bramwell and A. V. Hill, "The velocity of the pulse wave in man," Proc. R. Soc., London, Ser. B **93**, 298–306 (1922).
[17]J. Xu, "Do we need a better approach for measuring pulse-wave velocity?," Ultrasound Med. Biol. **29**, 1373–1374 (2003).
[18]M. Karamanoglu, "Errors in estimating propagation distances in pulse wave velocity," Hypertension **41**, E8–E8 (2003).
[19]N. Chubachi, H. Kanai, R. Murata, and Y. Koiwa, "Measurement of local pulse wave velocity in arteriosclerosis by ultrasonic Doppler method," Proc.-IEEE Ultrason. Symp. **1**, 1747–1750 (1994).
[20]X. M. Zhang, M. Fatemi, R. R. Kinnick, and J. F. Greenleaf, "Noncontact ultrasound stimulated optical vibrometry study of coupled vibration of arterial tubes in fluids," J. Acoust. Soc. Am. **113**, 1249–1257 (2003).
[21]X. Zhang, M. Fatemi, and J. F. Greenleaf, "Inverse method for estimating elastic modulus of arterial walls," Proc. SPIE **5368**, 226–235 (2004).
[22]X. Zhang, R. R. Kinnick, M. Fatemi, and J. F. Greenleaf, "Noninvasive method for estimating of complex elastic modulus of arterial vessels," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 642–652 (2005).
[23]Polytec Scanning Vibrometer, *Software Manual 7.0*, Polytec Pi, Inc., Auburn, 2001.
[24]X. Zhang, R. R. Kinnick, C. Pislaru, M. Fatemi, and J. F. Greenleaf, "In vivo noninvasive method for measuring local wave velocity in femoral arteries of pig," J. Acoust. Soc. Am. **118**, 2005–2005 (2005) (150th meeting of the Acoustical Society of America/NOISE-CON 2005, 4aBB3).

The spatial context of free-ranging Hawaiian spinner dolphins (*Stenella longirostris*) producing acoustic signals

Marc O. Lammers
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734, and Oceanwide Science Institute, P.O. Box 61692, Honolulu, Hawaii 96839

Michiel Schotten
Department of Marine Biology, Groningen University. P.O. Box 14, 9750 AA Haren, The Netherlands, Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734, and Oceanwide Science Institute, P.O. Box 61692, Honolulu, Hawaii 96839

Whitlow W. L. Au
Marine Mammal Research Program, Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

(Received 13 June 2005; revised 9 November 2005; accepted 22 November 2005)

To improve our understanding of how dolphins use acoustic signals in the wild, a three-hydrophone towed array was used to investigate the spatial occurrence of Hawaiian spinner dolphins (*Stenella longirostris*) relative to each other as they produced whistles, burst pulses, and echolocation clicks. Groups of approximately 30 to 60 animals were recorded while they traveled and socialized in nearshore waters off Oahu, Hawaii. Signaling animals were localized using time of arrival difference cues on the three channels. Sequences of whistles occurred between dolphins separated by significantly greater distances than animals producing burst pulses. Whistles typically originated from dolphins spaced widely apart (median = 23 m), supporting the hypothesis that whistles play a role in maintaining contact between animals in a dispersed group. Burst pulses, on the other hand, usually came from animals spaced closer to one another (median = 14 m), suggesting they function as a more intimate form of signaling between adjacent individuals. The spacing between echolocating animals was more variable and exhibited a bimodal distribution. Three quarters of echolocating animals were separated by 10 m or more, suggesting that the task of vigilance in a pod may not be shared equally by all members at all times. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151804]

PACS number(s): 43.80.Ka [JAS] Pages: 1244–1250

I. INTRODUCTION

Dolphins communicate using a combination of the visual, tactile, acoustic, and possibly chemosensory channels (Herman and Tavolga, 1980). Of these, only the acoustic modality allows for signaling over ranges greater than tens of meters. It can therefore be assumed that acoustic signals are the primary means by which delphinids mediate social processes that involve group coordinated behaviors (e.g., foraging and predator defense), navigation, and maintaining contact between widely dispersed individuals.

Dolphins produce a variety of sounds, which can be classified broadly into three categories: Echolocation click trains, burst pulse click trains, and whistles (Fig. 1). The clicks of most species are extremely short (~50 μs), broadband signals, extending in frequency from 2 to over 200 kHz (Au, 1993) and are emitted from the animal's forehead in a narrow directional beam (Au, 1993). The primary difference between echolocation clicks and burst pulse clicks is the number of clicks produced per unit time (Lammers *et al.*, 2004) and difference in amplitude (Au *et al.*, 1987). Echolocation clicks are used for sensing the surrounding environment, so they are generally emitted only after the echo of the previous click has been received by the dolphin, plus an additional period [15–45 ms (Au, 1993)] used to process the echo. Burst pulse signals are broadband click trains similar to those used in echolocation, but with interclick intervals of only a few (0.5–10) milliseconds (Lammers *et al.*, 2004). Because these intervals are considerably shorter than the processing period generally associated with echolocation, and because they are often recorded during periods of high social activity, burst pulse click trains are thought to instead play an important role in communication (Popper, 1980; Overstrom, 1983; Herzing, 1988).

In addition to clicks, many dolphin species produce long-duration frequency-modulated whistles, which are also associated with social communication (Herman and Tavolga, 1980), and which can be emitted simultaneously with burst pulse or echolocation clicks (Cranford, 2000). Whistles are among the most variable signals produced by dolphins. Perhaps the most common explanation for their function is that of "signature" signals that allow individuals to identify themselves to one another (Caldwell and Caldwell, 1965; Caldwell *et al.*, 1990; Tyack, 2000). The so-called "signature whistle hypothesis" has received support from numerous studies involving captive and restrained animals (Tyack, 1986; Caldwell *et al.*, 1990; Sayigh *et al.*, 1990; Janik *et al.*,

 0001-4966/2006/119(2)/1244/7/$22.50

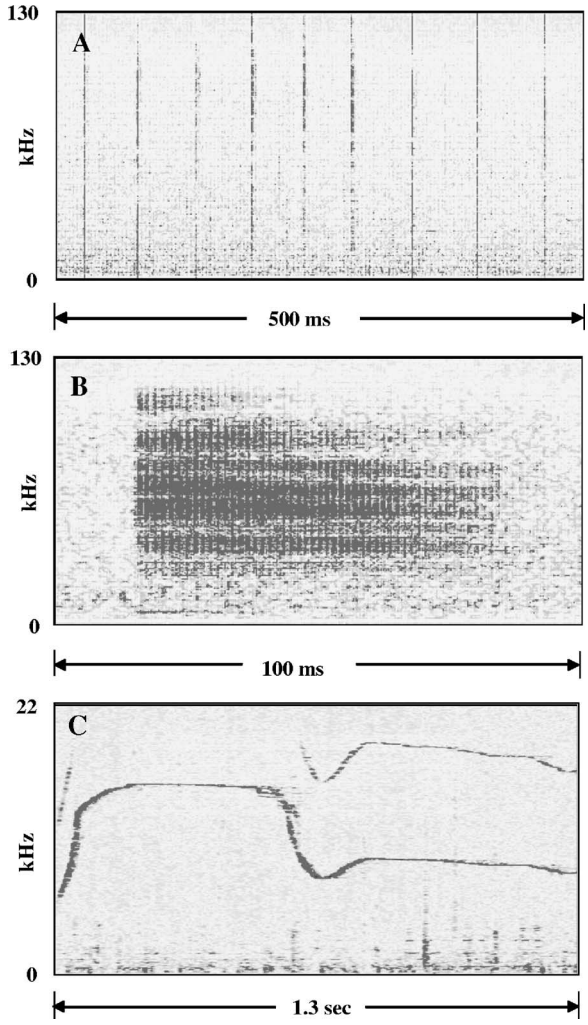


FIG. 1. Spectrogram examples of a dolphin echolocation click train (A), a burst pulse click train (B), and whistle harmonics (C).

1994; Janik and Slater, 1998), as well as from field studies of free-ranging animals (Smolker *et al.*, 1993; Herzing, 1996; Janik, 2000). Some, however, have argued that a simple signature function alone cannot account for the diversity of signals observed in socially interactive dolphin groups (McCowan and Reiss, 1995a, 2001), and that additional communicative functions are therefore likely also involved.

Dolphins produce whistles with fundamental frequencies that are usually in the human audible range (below 20 kHz). However, whistles typically also have harmonics, which occur at integer multiples of the fundamental and extend well beyond the range of human hearing (Lammers *et al.*, 2003). Harmonics are integral components of tonal signals produced by departures of the wave form from a sinusoidal signal. It has been proposed that this signal feature could function as a cue that allows listening animals to infer the orientation and direction of movement of a signaling dolphin (Miller, 2002; Lammers and Au, 2003). Whistles and their harmonics may therefore be important in mediating group cohesion and coordination.

Simple modeling of the dolphin's sound generator suggests that whistles have a wider beam pattern than clicks, especially at the fundamental and lower harmonic frequencies of the whistle (Au *et al.*, 1999; Lammers and Au, 2003). It is therefore typically presumed that whistles are perceived by listening dolphins over longer ranges than burst pulse clicks. Based on the hypothesis that dolphins use whistles to maintain group coordination when they are outside of each other's visual range, the prediction tested in the current study is that whistles are produced by dolphins that are spaced more widely apart from each other than dolphins exchanging burst pulse signals, which may function as a more "intimate" form of signaling between adjacent individuals. In addition, the spatial relationship is examined between dolphins that echolocated simultaneously to determine whether spacing between them is random, or whether it tends to follow patterns that could suggest how echolocation is used within a group of dolphins.

II. MATERIALS AND METHODS

To examine the spatial context of acoustic signaling in dolphins, signals were recorded from free-ranging Hawaiian spinner dolphins (*Stenella longirostris*) resident along the leeward coast of the island of Oahu, Hawaii. Spinner dolphins are a gregarious island-associated species commonly found along many of Hawaii's coastlines (Norris *et al.*, 1994). Their occurrence and daily behavioral cycle tend to follow predictable patterns, marked by periods of rest, travel, and socializing during different times of the day (Lammers, 2004). Their social and echolocation signals have been previously characterized by Lammers *et al.* (2003) and Schotten *et al.* (2004), respectively. Acoustic signals produced by spinner dolphins were recorded using a three-hydrophone line array towed behind a boat, as described by Lammers and Au (2003). A towed three-hydrophone line array enables one to localize dolphin signals in a two-dimensional (2D) plane by using the differences in time of signal arrival at the three hydrophones. This method can be used to record dolphins that are traveling in the same direction as the boat (since the array has to be towed), and generally to localize dolphins up to distances of approximately $12^{*}S$ away, where S represents the spacing between hydrophones.

A. Data collection

The hydrophone array used (Fig. 2) was composed of three ITC 1094A spherical omnidirectional hydrophones, with a calibrated sensitivity of approximately −205 dB re 1 V/μPa. The hydrophones were spaced 8 m apart from each other, with the cable attached to a 1.2 cm thick nylon line. A custom-made amplifier/line driver was coupled to each hydrophone, providing 40 dB of gain, as well as a 3 kHz high-pass filter to eliminate engine and water flow noise. At the front of the array, a 15×30 cm polyvinyl chloride (PVC) "towfish" with downward-angled wings and 6 kg of lead weight was used to sink the array to an operating depth of approximately 2 m. At the end of the array, a 3 m long 1.0 cm thick nylon "tattletale" line—with 20 cm long cable ties attached perpendicularly to the line 10 cm apart—was used to create drag and thus maintain tension on the array while being towed.

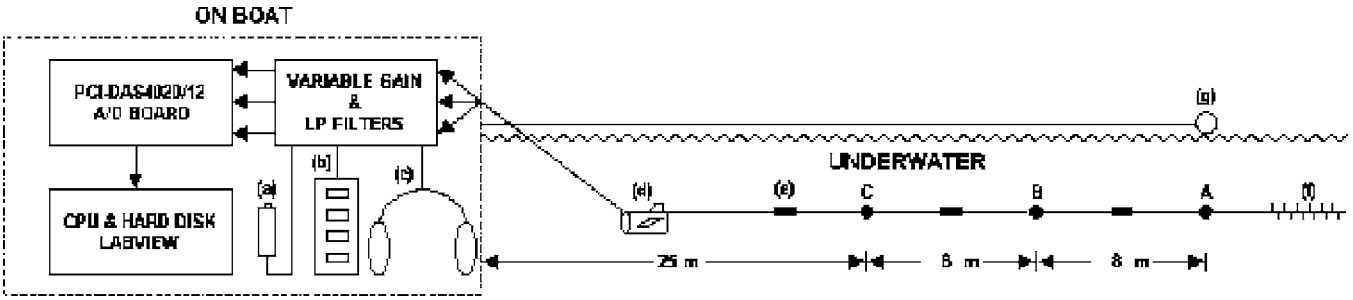


FIG. 2. Schematic of the three-hydrophone line array recording unit. The system is composed of a signal conditioning unit with low-pass filters and variable gain up to 35 dB, a four-channel PCI-DAS4020/12 A/D board operated from a lunchbox computer running custom-written LABVIEW6*i*™ data acquisition software, a manual TTL trigger (a) to initiate 10 s recording periods, and an LED meter (b) and headset (c) to detect and monitor the level of incoming dolphin signals. Underwater, a PVC tow-fish (d) sinks the array to a depth of approximately 2 m while being towed, and signals are collected on hydrophones A, B, and C, that are each coupled to amplifier/ line drivers (e) providing 40 dB of preamplification as well as 3 kHz high-pass filters. A tattletale (f) at the end maintains tension on the array during towing, and a 40 cm diameter buoy (g) above water marks hydrophone A for the observers on board (from Lammers and Au, 2003).

A custom-built signal conditioning unit on board the research vessel amplified and low-pass filtered the incoming signals from the three channels. There was the option of providing 35 dB of additional amplification in 5 dB steps, and programmable Lattice Semiconductors™ ispPAC80 fifth-order filter chips were used to low-pass filter the signals at either 50 kHz or 150 kHz, depending on the analog/digital (A/D) sample rate that was used, to avoid aliasing. During all recordings, the variable gain was set at 30 dB, providing a total of 70 dB of amplification. The middle hydrophone (Channel B) was used to detect the presence of dolphin signals, by means of both headphones (for signals that had energy in the human audio range) and a light-emitting diode (LED) meter (which could also detect signals that were purely ultrasonic). A four-channel simultaneous sampling Measurement Computing™ A/D converter board, the PCI-DAS4020/12, was used to digitize incoming signals on the three channels, at a sample rate of either 125,000 samples/s (for field recordings made in May and June 2003) or 400,000 samples/s (for recordings in November 2003) providing a Nyquist frequency of 62.5 and 200 kHz, respectively. This A/D board provided 12-bit resolution and had an input voltage of ±5 V. It was operated from a Pentium 850 MHz "lunchbox" computer, and the data acquisition and storage process was run by a custom-written LABVIEW6*i*™ program. Upon detection of dolphin signals, a 10 s sampling period was initiated by a manual trigger, and the resulting data files were automatically stored on the computer's hard drive.

The line array was towed from a 9.8 m boat powered by a 120 hp inboard diesel engine. A buoy towed separately and parallel to the array marked the position of the furthest hydrophone. Data collection was accomplished by either two or three persons, with one person operating the computer from the cabin below, one person driving the boat from the flying bridge, approximately 5 m above the water surface, and a third person who logged the presence of dolphins. In the absence of a third person, the boat operator logged the presence of dolphins. Radio contact was maintained between the bridge and the cabin, and when the computer operator detected dolphin signals and started a recording period, the data logger was notified to log the number and position of visible dolphins relative to the array. This was done to verify that dolphin positions calculated from the three-hydrophone localization algorithm were approximately correct.

Acoustic recordings from spinner dolphins were obtained along the leeward coast of Oahu, Hawaii, on six separate days in May, June, and November 2003. Dolphins were recorded when they traveled from their daytime resting areas in shallow waters to their evening offshore foraging grounds. To mitigate the potentially confounding effects of right-left ambiguity inherent in working with a line array, efforts were made to keep the boat positioned at one side and slightly ahead of all the dolphins in a pod, while traveling at approximately the same speed as the pod. If approached carefully, a group of dolphins could often be maintained in this relationship for several minutes. The array was also deployed in cases when dolphins appeared to be socially (and therefore, acoustically) active, but not traveling. The approach in those cases was to pass the group of dolphins on one side, while keeping the array in a straight line behind the boat (as indicated by the towed buoy). When the computer operator could no longer detect signals, the boat driver was notified to make a gradual 180° turn and, depending on the activity of the dolphins, drive by the group again. The dolphin pods recorded ranged in size from approximately 30 to 60 animals.

B. Data analysis

Recorded sounds were localized by evaluating the differences in time of sound arrival at each hydrophone using standard equations as described in Lammers and Au (2003). The speed of sound in water (c) used in the localization algorithm was 1533 m/s, which was calculated assuming a typical water temperature of 24.5 °C and a salinity of 34.9 ppt (Urick, 1983). The system's localization accuracy was calibrated by using an artificial omnidirectional sound source that produced dolphinlike frequency-modulated signals, which was placed at different angles and distances away from the center hydrophone up to a maximum of 30 m, while the array was kept stationary at a depth of 2 m. In all cases, localizations proved to be accurate to within 1–2 m from the actual position of the sound source.

Field recordings were initially inspected for the presence

of dolphin whistles and burst pulse signals of sufficient signal-to-noise ratio on all three channels, and accordingly categorized using COOL EDIT 96™ software. Subsequently, whistles were analyzed using a custom-written MATLAB 5.1™ program that implemented the three-hydrophone localization algorithm. Differences in time of whistle arrival at each hydrophone were obtained by cross correlating a particular whistle in the data files from Channels A and C with the same whistle in the file from Channel B. The largest peak of the resulting cross-correlation vector was used to establish the time of arrival difference between Channels A and B and between B and C. Since acoustic reflections from the water surface and poor signal-to-noise ratio can result in ambiguous or low cross-correlation results, recordings with those attributes were not considered for further analysis in order to maintain a high degree of confidence in the localizations.

For the analysis of burst pulse and echolocation clicks, arrival times had to be measured directly from the original data files. Wave forms were compared among the three channels to measure time of arrival differences between the three hydrophones. This was done by inspecting the first (or last) click of a particular burst pulse signal or echolocation click train on each channel, and logging the sample value of the first peak in amplitude of that click. The three logged sample values for a particular click were then entered in another custom-written MATLAB program to calculate the 2D position of the sound-emitting dolphin, using the localization algorithm. For each click train (both echolocation and burst pulse), this was repeated for up to several clicks, to ascertain that the localization for that train was correct. The x and y coordinates for each signal were then logged together with the time separation (Δt) between the localized signal and the signal produced just prior to it.

Finally, to establish distances between dolphins producing acoustic signals, it is necessary to determine whether the signal originated from the same dolphin or from a different dolphin than the signal immediately preceding and/or following it. For both whistles and burst pulses, this was accomplished by considering both their spatial and temporal relationships, as inferred from the calculated x,y coordinates for each signal, in combination with their measured time separation Δt, as well as the presumed maximum swimming speed. Bottlenose dolphins (*Tursiops truncatus*) can easily sustain a swimming speed of 2 m/s (Williams *et al.*, 1992), and have been reported capable of sustaining speeds over 8 m/s (Lang and Pryor, 1966). The smaller harbor porpoises (*Phocoena phocoena*) have reported mean swimming speeds of around 1 m/s and a maximum reported speed of 4.3 m/s (Otani, 2000). Therefore, in this study, it was assumed that spinner dolphins—which are intermediate in size compared to these other two species—do not swim faster than 8 m/s. Using the Pythagorean theorem, the distance d between two subsequent Signals A and B was calculated from their x,y coordinates as $d=\sqrt{(x_A-x_B)^2+(y_A-y_B)^2}$, with a ± 1 m accuracy. If $d/\Delta t \geqslant 8$ m/s (where Δt is the measured time separation between Signals A and B), it was assumed that Signals A and B originated from two different dolphins and the distance d was used for subsequent statistical analysis. This was a conservative criterion, since d and, therefore, $d/\Delta t$ are likely to be underestimates.

Echolocation click trains were often considerably longer in duration than whistles and burst pulses (up to several seconds), allowing time for spatial relationships to potentially change during the course of a click train. Therefore, to examine the spatial relationship between echolocating individuals, only click trains that overlapped in time with each other were considered. This ensured that separate animals were actually involved and that their spatial relationship did not change appreciably during the period examined.

III. RESULTS

The spatial context of whistle and burst pulse production

A total of 185 whistles, 172 burst pulse signals, and 94 echolocation click trains were localized. Signals that met the conditions mentioned previously were selected, and the distance d between the two dolphins that were assumed to have emitted the recorded signals was calculated for each of these exchanges. Distances were calculated for 27 burst pulse sequences, (5 of which represent average distances, for repeated signal sequences produced by presumably the same 2 animals) and 41 whistle sequences. Distances between animals echolocating concurrently were established in 44 cases. Examples of localizations for each type of signal are represented in Fig. 3, and the distributions of distances in each case are presented as histograms in Fig. 4. The distributions for both whistles and burst pulses were somewhat skewed toward shorter distances [Figs. 4(a) and 4(b)]. The minimum distance d was 4 m for whistle sequences and 3 m for burst pulse sequences, while the maximum calculated distance was 113 m for whistle sequences and 72 m for burst pulse sequences. Median distances were 23 m and 14 m for whistle and burst pulse sequences, respectively. Overall, the distance d between two dolphins was significantly larger for whistle sequences than for burst pulse sequences ($p=0.0054$, Mann–Whitney test, two tailed).

The distribution of distances for concurrently echolocating animals was distinctly bimodal [Fig. 4(c)]. The median distance between animals was 24 m, but two peaks occurred centered around 10 m and 40 m. In addition, 34 of the 44 instances (77%) were localized to animals separated by a distance greater than 10 m. Therefore, concurrent echolocation occurred primarily between individuals separated by several body lengths.

A. The temporal context of whistle and burst pulse sequences

In addition to the spatial context, the temporal context of dolphin signal sequences was also considered for whistles and burst pulses. Specifically, the timing of signals was investigated to determine whether any relationship exists with the spatial separation between dolphins. For each signal sequence, the calculated spatial separation d was plotted as a function of the temporal separation Δt between the two signals, as presented in Fig. 5. For whistle sequences [Fig. 5(a)], a very weak positive linear relationship was found (R^2

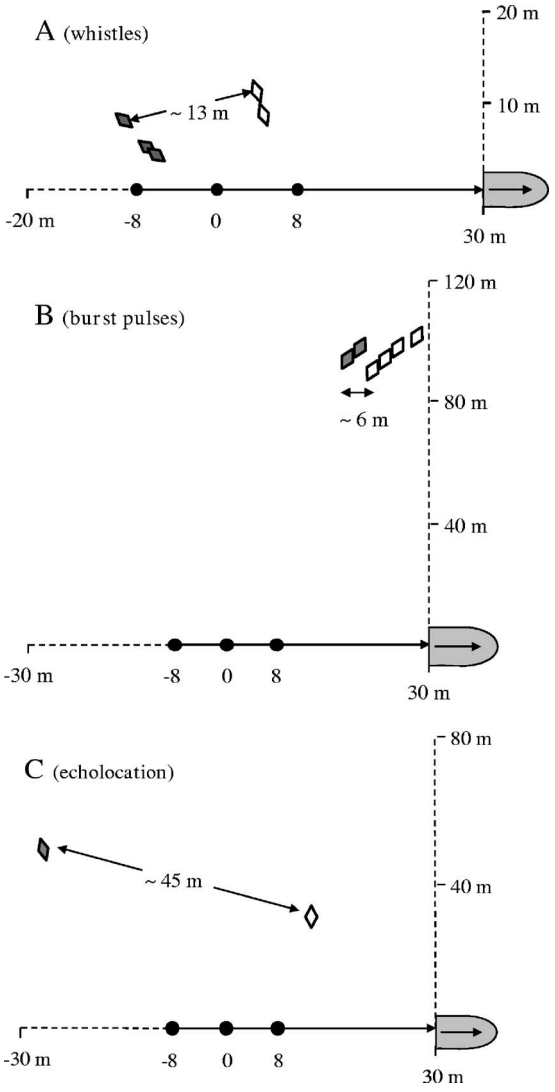


FIG. 3. Representative examples of localizations obtained for a whistle exchange (A), burst pulses exchange (B), and concurrent echolocation by two dolphins (C). The grey and white diamonds represent the successive positions of the two animals involved in the exchange, relative to the towed array and towing vessel. The black circles represent the hydrophones of the array, with the center hydrophone serving as the origin in the coordinate axes used to plot localizations.

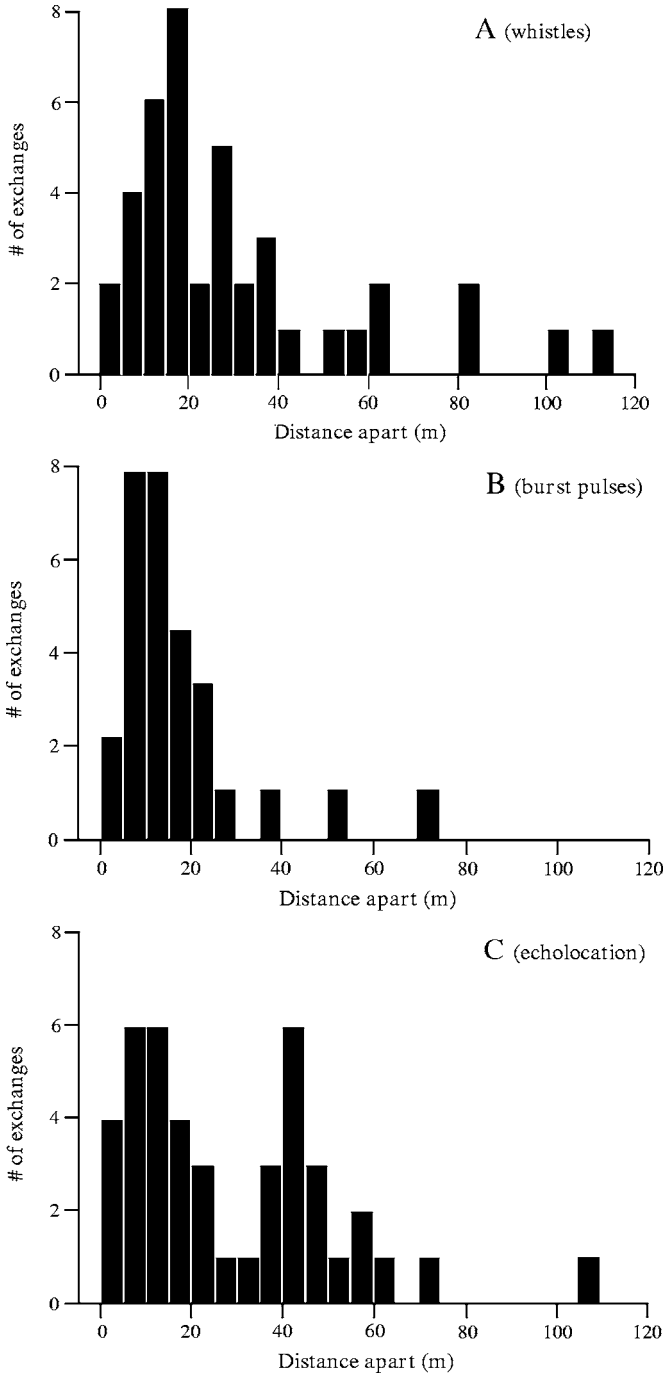


FIG. 4. Distributions of calculated separation distance between two dolphins for whistle sequences (A, $N=41$), burst pulse sequences (B, $N=27$), and concurrent echolocation click trains (C, $N=44$).

$=0.05$), however, this relationship was not significant ($P=0.148$). For burst pulse sequences on the other hand [Fig. 5(b)], a slightly stronger positive linear relationship was found between spatial and temporal separation of the signals ($R^2=0.21$) and this relationship proved to be significant ($P=0.013$). Therefore, burst pulse signals that were produced between two dolphins that were closer to each other were found to also follow each other more quickly in time, while this was not the case for whistle sequences.

IV. DISCUSSION

The results reveal that the production of whistles among spinner dolphins tends to take place between individuals separated by significantly greater distances than between animals producing burst pulses. It cannot be established at this point whether either whistles or burst pulses were "exchanged" between individual dolphins, since the term implies that information is transmitted from one animal to another, which then replies with a signal back to the first animal. However, the significant difference in spatial separation between dolphins emitting whistles versus those emitting burst pulses, does suggest that these signals have different functions which are directly related to the dolphins' proximity to one another, thus suggesting communicative exchanges. Additionally, evidence of whistle matching by wild bottlenose dolphins (*Tursiops truncatus*) suggests that dolphins, in fact, do address one another acoustically (Janik, 2000), leaving open the possibility that the sequences observed do represent exchanges of signals.

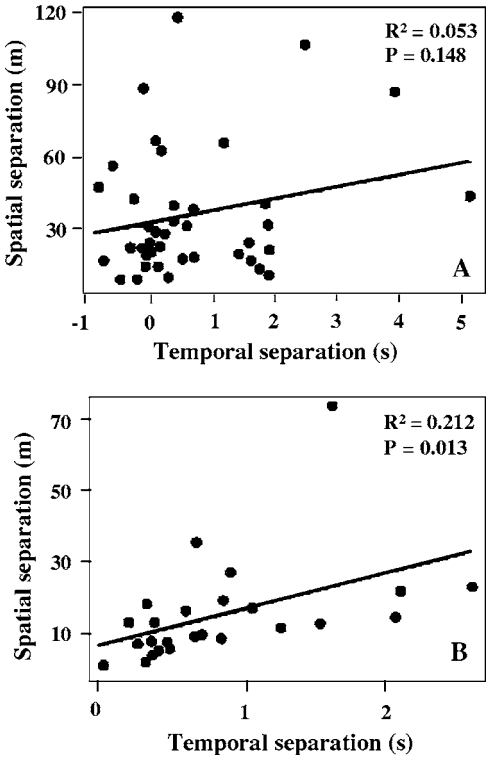


FIG. 5. Calculated separation distances plotted as a function of the measured separation in time between two signals, for whistle sequences (A) and burst pulse sequences (B), with least-square linear regression lines through the data points.

Our findings provide empirical evidence that whistles and burst pulses play distinct functional roles. Specifically, most whistling appears to take place near or beyond the periphery of visual contact, whereas burst pulsing typically occurs within it (assuming a visual range of ~20 m in local inshore waters). Whistles, therefore, are likely important in signaling between individuals dispersed in a pod (Fig. 6).This supports the hypothesis that whistles, and the acoustic cues present in them, play an important role in mediating group behavior and cohesion (Norris *et al.*, 1994; Lammers and Au, 2003). Janik and Slater (1998) previously found evidence for this through acoustic observations of a group of captive dolphins, and Smolker *et al.* (1993) documented the use of contact calls between mothers and calves, but this is the first known evidence from free-ranging pods.

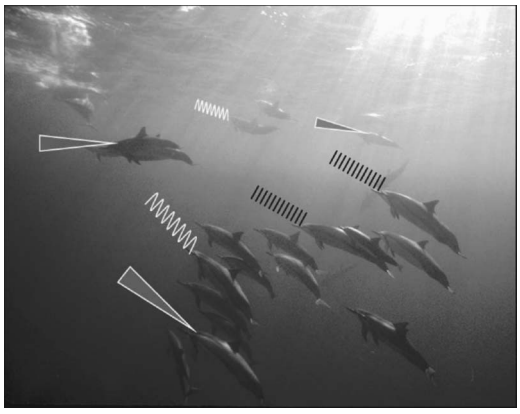


FIG. 6. A hypothetical example of the proposed spatial occurrence of acoustic signals in a spinner dolphin pod. The white sinusoids represent whistle production, the black bars represent the production of burst pulses, and the outlined gray cones are the occurrence of echolocation click trains (photo courtesy of Andre Seale).

Burst pulsing, on the other hand, takes place primarily between nearby animals in a group (Fig. 6). This is consistent with their hypothesized function as "emotive" signals, as several researchers have previously suggested (Lilly and Miller, 1961; Herzing, 1988, 1996; Bloomqvist and Amundin, 2004). Burst pulses have been implicated primarily in aggressive or agonistic contexts, such as confrontational "head-to-head" behaviors between individuals (Caldwell and Caldwell, 1967; Overstrom, 1983; McCowan and Reiss, 1995b; Bloomqvist and Amundin, 2004). However, contexts ranging from courtship behaviors to alarm responses have also been associated with burst pulsing (Herzing, 1996). Their occurrence, primarily among closely spaced individuals, is consistent with the task of communicating emotively charged information to potential mates or affiliated individuals, or to perhaps support visual displays. In addition, by presumably being more directional than whistles, burst pulses may function as more intimate forms of signaling than whistles, and be meant primarily for individuals in the immediate vicinity.

The bimodal distance distribution of concurrently echolocating dolphins is interesting and potentially revealing of how echolocation is used in a pod. Specifically, the fact that over three-quarters of cases were of animals separated by 10 m or more suggests that the task of vigilance in a pod may not be shared equally by all members simultaneously. Spinner dolphins rarely swim alone for any length of time, but rather occur in pairs or small clusters of several animals that are usually separated by less than one body length from each other (Norris *et al.*, 1994). It is therefore somewhat surprising that very little concurrent echolocation was observed between closely spaced individuals. Rather, the first peak—occurring around 10 m—is consistent with the approximate horizontal diameter of a spinner dolphin subpod (as personally observed by one of the authors—M. L.). Therefore, the results suggest that vigilance through echolocation may be maintained by animals strategically located in a subpod rather than by all individuals echolocating at random (Fig. 6). Furthermore, the second peak at 40 m likely represents echolocating animals in adjacent subpods.

Finally, the timing of signals relative to the physical separation of the signalers was examined to determine whether they are related. It has been suggested that spinner dolphin chorusing during certain times of the day may be important in the process of coordinating group behavior (Norris *et al.*, 1994). The timing of signals might play a role in this process by cueing nearby dolphins on the physical proximity of the signalers to one another, and thereby communicate their general behavioral disposition (resting, traveling, socializing, etc.). As an example, signals closely timed (or overlapping) might be indicative of close proximity, which is characteristic of resting and/or social behavior. Conversely, widely spaced signalers (together with widely timed signals) could represent a traveling pod, or one not very synchronized. For whistles, however, no such correlation was

noted between separation and timing. Burst pulses, on the other hand, did exhibit a weak but significant correlation, and followed each other more quickly in time when they were exchanged by two dolphins that were closer to one another. While perhaps not sufficient to warrant assuming a direct relationship, this result does suggest that a more detailed examination of the timing of burst pulses within various behavioral contexts could be fruitful.

ACKNOWLEDGMENTS

The authors would like to thank Lisa Albinsson, Lisa Davis, Mark Deakos, and Dieter Lammers for their dedicated assistance with data collection in the field, and Andre Seale for the photo in Fig. 6. Also, Paul Nachtigall provided important logistical support and valuable commentary, and Wim Wolff and John Videler reviewed an earlier draft of this manuscript. Two anonymous reviewers also provided helpful comments. An anonymous foundation, the University of Hawaii Sea Grant Program, the Hawaii State Department of Transportation, and Ko'Olina Marina provided financial and/or logistical support for the project. The Royal Netherlands Academy of Arts and Sciences (KNAW; Grant No. SBP/JK/2002-19) and the Royal Dutch Society of Sciences (KHMW) provided financial support for one of the authors (M. S.). The project was conducted under NMFS Permit No. 1000-1617. This is HIMB Contribution No. 1215.

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

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). "Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L., Penner, R. H., and Turl, C. W. (**1987**). "Propagation of beluga echolocation signal," J. Acoust. Soc. Am. **82**, 807–813.

Bloomqvist, C., and Amundin, M. (**2004**). "High-frequency burst-pulse sounds in agonistic/aggressive interactions in bottlenose dolphins (*Tursiops truncatus*)," in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 425–431.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "A review of the signature whistle hypothesis for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego, CA), pp. 199–234.

Caldwell, M. C., and Caldwell, D. K. (**1967**). "Intraspecific transfer of information via the pulsed sound in captive Odontocete Cetaceans," in *Animal Sonar Systems: Biology and Bionics*, edited by R. G. Busnel (Laboratoire de Physiologie Acoustic: Jouy-en-Josas, France), pp. 879–936.

Caldwell, M. C., and Caldwell, D. K. (**1965**). "Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*)," Nature (London) **207**, 434–435.

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

Herman, L. M., and Tavolga, W. N. (**1980**). "The communication systems of cetaceans," in *Cetacean Behaviour: Mechanisms and Function*, edited by L. M. Herman (Wiley-Interscience, New York), pp. 149–209.

Herzing, D. L. (1988). "A quantitative description and behavioral association of a burst-pulsed sound, the squawk, in captive bottlenose dolphins, *Tursiops truncatus*," M.S. thesis, San Francisco State University, CA.

Herzing, D. L. (**1996**). "Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*," Aquat. Mamm. **22**, 61–79.

Janik, V. M. (**2000**). "Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)," Science **289**, 1355–1357.

Janik, V. M., and Slater, P. J. B. (**1998**). "Context specific use suggests that bottlenose dolphin signature whistles are cohesion calls," Anim. Behav. **56**, 829–838.

Janik, V. M., Denhardt, G., and Dietmar, T. (**1994**). "Signature whistle variation in a bottlenosed dolphin, *Tursiops truncatus*," Behav. Ecol. Sociobiol. **35**, 243–248.

Lammers, M. O. (**2004**). "Occurrence and behavior of Hawaiian spinner dolphins (*Stenella longirostris*) along Oahu's leeward and south shores," Aquat. Mamm. **30**, 237–250.

Lammers, M. O., Au, W. W. L., Aubauer, R., and Nachtigall, P. E. (**2004**). "A comparative analysis of echolocation and burst-pulse click trains in *Stenella longirostris*," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. M. Vater (University of Chicago Press, Chicago), pp. 414–419.

Lammers, M. O., and Au, W. W. L. (**2003**). "Directionality in the whistles of Hawaiian spinner dolphins (*Stenella longirostris*): A signal feature to cue direction of movement?" Marine Mammal Sci. **19**, 249–264.

Lammers, M. O., Au, W. W. L., and Herzing, D. L. (**2003**). "The broadband social acoustic signaling behavior of spinner and spotted dolphins," J. Acoust. Soc. Am. **114**, 1629–1639.

Lang, T. G., and Pryor, K. (**1966**). Science **152**, 531–533.

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

McCowan, B., and Reiss, D. (**2001**). "The fallacy of 'signature whistles' in bottlenose dolphins: A comparative perspective of 'signature information' in animal vocalizations," Anim. Behav. **62**, 1151–1162.

McCowan, B., and Reiss, D. (**1995a**). "Quantitative comparison of whistle repertoires from captive adult bottlenose dolphins (Delphinidae, *Tursiops truncatus*): A re-evaluation of the signature whistle hypothesis," Ethology **100**, 194–209.

McCowan, B., and Reiss, D. (**1995b**). "Maternal aggressive contact vocalizations in captive bottlenose dolphins (*Tursiops truncatus*): Wide-band, low-frequency signals during mother/aunt-infant interactions," Zoo Biol. **14**, 293–309.

Miller, P. J. O. (**2002**). "Mixed-directionality of killer whale stereotyped calls: A direction-of-movement cue?" Behav. Ecol. Sociobiol. **52**, 262–270.

Norris, K. S., Würsig, B., Wells, R. S., and Würsig, M. (**1994**). *The Hawaiian Spinner Dolphin* (University of California Press, Berkeley).

Otani, S. (**2000**). "Diving behavior and swimming speed of a free-ranging harbor porpoise, *Phocoena phocoena*," Marine Mammal Sci. **16**, 811–814.

Overstrom, N. A. (**1983**). "Association between burst-pulse sounds and aggressive behavior in captive Atlantic bottlenose dolphins (*Tursiops truncatus*)," Zoo Biol. **2**, 93–103.

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

Sayigh, L. S., Tyack, P. L., Wells, R. S., Scott, M. D., and Irvine, A. B. (**1990**). "Signature differences in signature whistles production of free-ranging bottlenose dolphins *Tursiops truncatus*," Behav. Ecol. Sociobiol. **36**, 171–177.

Schotten, M., Au, W. W. L., Lammers, M. O., and Aubauer, R. (**2004**). "Echolocation recordings and localization of wild spinner dolphins (*Stenella longirostris*) and pantropical spotted dolphins (*S. attenuata*) using a four hydrophone array," in: *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. M. Vater (University of Chicago Press, Chicago), pp. 393–400.

Smolker, R. A., Mann, J., and Smuts, B. B. (**1993**). "Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants," Behav. Ecol. Sociobiol. **33**, 393–402.

Tyack, P. L. (**2000**). "Dolphins whistle a signature tune," Science **289**, 1310–1311.

Tyack, P. L. (**1986**). "Whistle repertoires of two bottlenose dolphins, *Tursiops truncatus*: Mimicricy of signature whistles?" Behav. Ecol. Sociobiol. **18**, 251–257.

Urick, R. J. (**1983**). *Principles of Underwater Sound*. 3rd Ed. (McGraw–Hill, New York).

Williams, T. M., Friedl, W. A., Fong, M. L., Yamada, R. M., Sedivy, P., and Haun, J. E. (**1992**). "Travel at low energetic cost by swimming and wave-riding bottlenose dolphins," Nature (London) **355**, 821–823.

Phase effects in masking by harmonic complexes in birds[a)]

Amanda M. Lauer[b)] and Robert J. Dooling
Department of Psychology, University of Maryland, College Park, Maryland 20742 and Center for the Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

Marjorie R. Leek
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307-5001 and Center for the Comparative and Evolutionary Biology of Hearing, University of Maryland, College Park, Maryland 20742

Jennifer J. Lentz
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 3 May 2005; revised 14 October 2005; accepted 23 November 2005)

Masking by harmonic complexes depends on the frequency content of the masker and its phase spectrum. Harmonic complexes created with negative Schroeder phases (component phases decreasing with increasing frequency) produce more masking than those with positive Schroeder phases (increasing phase) in humans, but not in birds. The masking differences in humans have been attributed to interactions between the masker phase spectrum and the phase characteristic of the basilar membrane. In birds, the similarity in masking by positive and negative Schroeder maskers, and reduced masking by cosine-phase maskers (constant phase), suggests a phase characteristic that does not change much along the basilar papilla. To evaluate this possibility, the rate of phase change across masker bandwidth was varied by systematically altering the Schroeder algorithm. Humans and three species of birds detected tones added in phase to a single component of a harmonic complex. As observed in earlier studies, the minimum amount of masking in humans occurred for positive phase gradients. However, minimum masking in birds occurred for a shallow negative phase gradient. These results suggest a cochlear delay in birds that is reduced compared to that found in humans, probably related to the shorter avian basilar epithelia. [DOI: 10.1121/1.2151816]

PACS number(s): 43.80.Lb, 43.66.Gf [JAS] Pages: 1251–1259

I. INTRODUCTION

While the representation of the spectral characteristics of sound along the avian basilar papilla is fairly well understood, less is known about how temporal information is processed along the papilla. Yet, the acoustic communication system of birds involves some of the most temporally complex acoustic signals in nature and it is becoming increasingly clear that birds perceive much of this complexity. In mammals, there is considerable evidence of an interaction between the spectral and temporal characteristics of a sound waveform and the response characteristics of the basilar membrane (e.g., Reccio and Rhode 2000). The temporal waveform shape may also influence the internal representation of sound on the bird basilar papilla, but this interaction between stimulus and response characteristics of the papilla has been less well explored. The aim of the present study was to examine how temporal response properties of the avian basilar papilla interact with the waveform shape and spectral characteristics of complex sounds using behavioral masking methods. Masking of a probe tone was measured for harmonic complex maskers with varying phase spectra. Component phases were selected to provide systematic changes in the temporal waveforms of the maskers without altering the long-term amplitude spectrum.

Variations in the phase spectrum of a harmonic complex sound may produce large differences in waveform shape and in the temporal pattern of instantaneous frequencies within the waveforms. The maskers used in this study were constructed with harmonic component starting phases selected according to a modification of an algorithm developed by Schroeder (1970)

$$\theta_n = C\pi n(n+1)/N, \tag{1}$$

where θ_n represents the phase of the nth harmonic, N is the total number of harmonics, and C is a scalar (Lentz and Leek 2001). Phase spectra for several of the complexes are depicted in Fig. 1(a). The Schroeder-phase stimuli feature monotonic changes in phase across frequency that produce upward or downward sweeps in instantaneous frequency within each period of the complex, as illustrated in Fig. 1(b). The direction of frequency sweep may be reversed by altering the sign of the Schroeder algorithm [left versus right columns in Fig. 1(b)]. The speed of frequency sweep may be manipulated by the choice of constant scalar in the phase selection algorithm (Lentz and Leek 2001). Complexes made with the same scalar value, but opposite sign, result in time

[a)]Portions of this work were presented at the 141st Meeting of the Acoustical Society of America, Chicago, IL, June 2001 and the 6th International Congress on Neuroethology, Bonn, Germany, August 2001.
[b)]Electronic mail: alauer@psyc.umd.edu

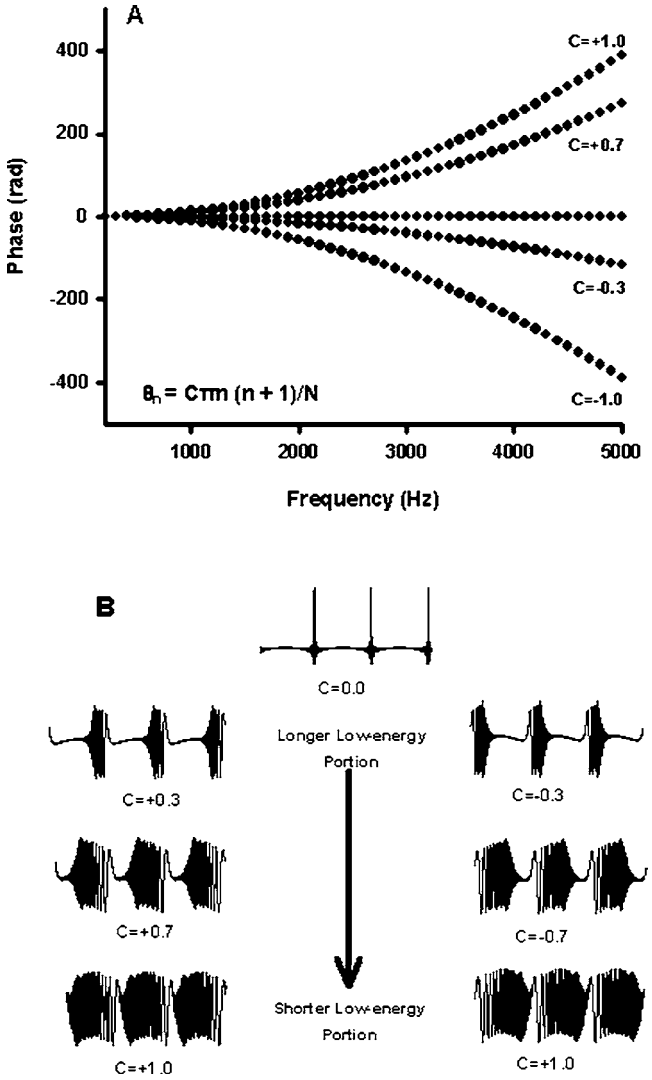


FIG. 1. Phase spectra (a) and temporal waveforms (b) for several of the harmonic complex maskers. Each waveform is constructed of equal-amplitude cosine tones from 200 to 5000 Hz, with a fundamental frequency of 100 Hz. Phases were selected according to scaled modifications of the algorithm shown in the lower left corner of (a) (θ_n= the nth harmonic, N =total number of harmonics, and C=a scalar). Three periods (30 ms) are shown for each temporal waveform.

reversed waveforms, with phases increasing (positive Schroeder) or decreasing (negative Schroeder) and instantaneous frequency decreasing or increasing, respectively.

Because none of these phase manipulations has any effect on the long-term frequency spectra, a theory of masking based entirely on the spectrum would predict equal masking from all of these Schroeder complexes. However, the original negative and positive Schroeder-phase maskers (i.e., those with a scalar C of ±1.0) can produce large masking differences in humans (Smith *et al.* 1986; Kohlrausch and Sander 1995; Carlyon and Datta 1997a; 1997b; Summers and Leek 1998; Lentz and Leek 2001; Oxenham and Dau 2001). This effect has been attributed to cochlear processing mechanisms that may interact with the stimulus phase spectrum to produce an altered "internal" waveform. Kohlrausch and Sander (1995) proposed that the mammalian basilar membrane alters the shape of the positive Schroeder-phase masker such that the corresponding internal waveform is more modulated than the internal waveform corresponding to the negative Schroeder-phase masker. This increased modulation creates portions of low energy within each period in the positive Schroeder internal waveform within which a signal tone may be more easily detected in a masking task. Negative Schroeder-phase maskers do not undergo such alteration in internal shape, and therefore masking is greater. Hearing-impaired listeners do not show these large masking differences between the positive and negative waveforms (Summers and Leek 1998). This effect is thought to be related to the loss of the nonlinear active processing mechanism in a damaged cochlea.

Recent studies have demonstrated dramatic differences in masking by Schroeder-phase harmonic complexes between birds and humans when the scalar is ±1.0. In contrast to humans, the negative-phase and positive-phase waveforms produce similar amounts of masking of a 2.8 kHz tone in three species of birds: budgerigars, zebra finches, and canaries (Dooling *et al.* 2001; Leek *et al.* 2000). The differences in the patterns of masking in birds and humans have been attributed to structural differences in the mammalian and avian ears. The length of the basilar papilla in birds is an order of magnitude smaller than the human cochlea (Gleich *et al.* 1994; Manley *et al.* 1993). In addition, the stiffness gradient along the basilar papilla is steeper in the avian ear (von Bekesy 1960). Finally, the frequency-dependent cochlear delay is much shorter in starlings and pigeons than in most mammals (Gleich and Narins 1988). These anatomical and physiological differences undoubtedly affect the nature of the traveling wave in birds.

To further investigate differences in phase processing by the inner ears in birds and humans, we used harmonic complex maskers with phase spectra selected according to these scaled modifications of the Schroeder algorithm to produce waveforms with envelopes that vary systematically between highly peaky (cosine phase) and very flat. Both positive-phase and negative-phase maskers were tested. Maskers with more modulated envelopes are characterized by faster repeating frequency sweeps (once each period) and longer low energy portions, and maskers with flatter envelopes contain slower repeating frequency sweeps and very short low energy portions. The internal phase characteristic of the basilar membrane (or basilar papilla) in the region of maximum displacement for a particular frequency may be estimated by finding the least effective scaled Schroeder-phase masker (Lentz and Leek 2001). The phase spectrum of this masker is assumed to approximately cancel the phase characteristic of the cochlea in the frequency region of the signal, thereby creating an internal within-channel waveform with a highly peaked shape (effectively, a cosine-phase internal waveform). In this way, psychophysical measures have been used to estimate the phase characteristic of the human basilar membrane by determining the least effective masker in a set of scaled Schroeder-phase stimuli. In humans, the least effective masker varies somewhat across listeners and across frequency, but is always a positive-phase masker for mid- to high-frequency signals (Lentz and Leek 2001). These results

in humans suggest that the internal waveforms corresponding to the positive-phase maskers are more modulated, enabling a probe tone to be more easily detected.

In this study, the phase response of the basilar papilla in birds was investigated in order to infer temporal characteristics of the traveling wave. Budgerigars, zebra finches, and canaries were tested in a masking paradigm using stimuli similar to those used for testing humans (Lentz and Leek 2001; Oxenham and Dau 2001). In addition, as a control, three humans were tested using the same methodologies used in testing birds. Thresholds for detecting tones embedded in waveforms with systematically varying shape were measured using the Method of Constant Stimuli and operant conditioning techniques.

II. METHODS

A. Subjects

Three adult zebra finches, three adult budgerigars, and three adult canaries were used as subjects. The birds were kept on a normal day/night cycle correlated with the season and maintained at approximately 90% of their free-feeding weights. For comparison, three young adult humans (laboratory staff members) also participated in the experiment. All birds and humans had hearing within normal limits for their species, as shown by their audiograms. Animal housing and care met all standards of the University of Maryland Animal Care and Use Committee (ACUC), College Park, MD. All research was approved by the ACUC and the Internal Review Board.

B. Stimuli

Stimuli were harmonic maskers and maskers plus a signal tone. The masking stimuli were constructed by summing equal-amplitude tones from 200 to 5000 Hz, with a fundamental frequency of 100 Hz. The phases of the tones were selected according to a modification of the Schroeder-phase algorithm [Eq. (1)] described earlier. Maskers were generated for scalars (C) ranging between −1.0 and +1.0. These two end-value scalars result in the original Schroeder-negative and Schroeder-positive phase maskers used in earlier masking studies. When $C=0.0$, a highly modulated cosine-phase waveform is produced, characterized by a very high peak once each period with low-amplitude energy during the rest of the period. Negatively valued scalar stimuli have a rising frequency sweep within the masker period, and positively valued scalars have a falling frequency sweep within each period. Changing the scalar changes the rate of the frequency sweep, in that scalars closer to zero produce more rapid frequency sweeps than those close to ±1.0. Seven examples of these masker waveforms are depicted in Figure 1(b). The different scalars generate maskers on a continuum of both frequency sweep rate within each period and relative proportion of low versus high energy within each period.

Twenty-one scalars were tested for zebra finches and 13 scalars were tested for budgerigars and canaries. The maskers were 260 ms in duration with 20 ms raised-cosine rise/fall times. The maskers were presented at a level of 80 dB SPL (63 dB SPL per harmonic component). The signal was a 2.8 kHz tone added in phase to the 2.8 kHz masker component. The duration of the signal was the same as the masker, including the rise and fall times.

Stimuli were created digitally by summing waveforms at the component frequencies with the appropriate phases and amplitudes. All stimuli were created off line and stored as files for playback during the experiment. The sampling rate was 40 kHz. Each set of stimuli included a masker waveform alone, and a number of masker-plus-signal waveforms, at several signal-to-masker ratios. The signal level at threshold is reported as the level of the signal component added to the masker in decibels (dB) relative to the level of each masker component.

C. Testing Apparatus

The birds were tested in a wire cage ($23\times25\times16$ cm) mounted in a sound-isolation chamber (Industrial Acoustics Company, Bronx, NY, IAC-3). A response panel consisting of two microswitches with light-emitting diodes (LEDs) was mounted on the wall of the test cage just above the food hopper. Microswitches were tripped when a bird pecked the attached LED. The left microswitch served as the observation key, and the right microswitch served as the report key. During test sessions, the behavior of the animal was monitored by a video camera system (Sony HVM-322).

Test sessions were controlled by a computer (IBM Pentium III). The digital stimuli were output to a KEF loudspeaker (model 80C) via Tucker-Davis modules at masker levels of 80 dB SPL. Stimulus calibration was performed using a Larson-Davis System 824 sound level meter (model 824). Stimulus intensities were measured with a $\frac{1}{2}$-in. microphone attached to the sound level meter via a 3-m extension cable. The microphone was placed in front of the keys in the approximate position occupied by the bird's head during testing. Masker intensities were measured several times during the experiment to ensure that stimulus levels remained constant and the entire system was calibrated.

D. Training and testing procedures

Birds were trained by standard operant auto-shaping procedures (Dooling and Okanoya 1995) to peck two keys constructed of LEDs attached to microswitches. Birds pecked at the left key (observation key) during a repeating background for a random amount of time between 2 and 7 s until a target stimulus was alternated with the background sound. If the bird pecked the right key (report key) within 2 s of this alternating pattern, it received access to food from a hopper for 1 or 2 s. The dependent variable was percent correct on trials involving an alternating sound pattern. Failure to peck the report key within 2 s of the alternating pattern was recorded as a miss, and a new trial sequence was initiated. Thirty percent of all trials were sham trials in which the target sound was the same as the background sound. A peck to the report key during a sham trial was recorded as a false alarm, and the test chamber lights were extinguished for 5–15 s. Birds typically exhibited false alarm rates between 3 and 10%. Average false alarm rates were 3.50% for budgerigars, 4.77% for zebra finches, and 2.82% for canaries. Data

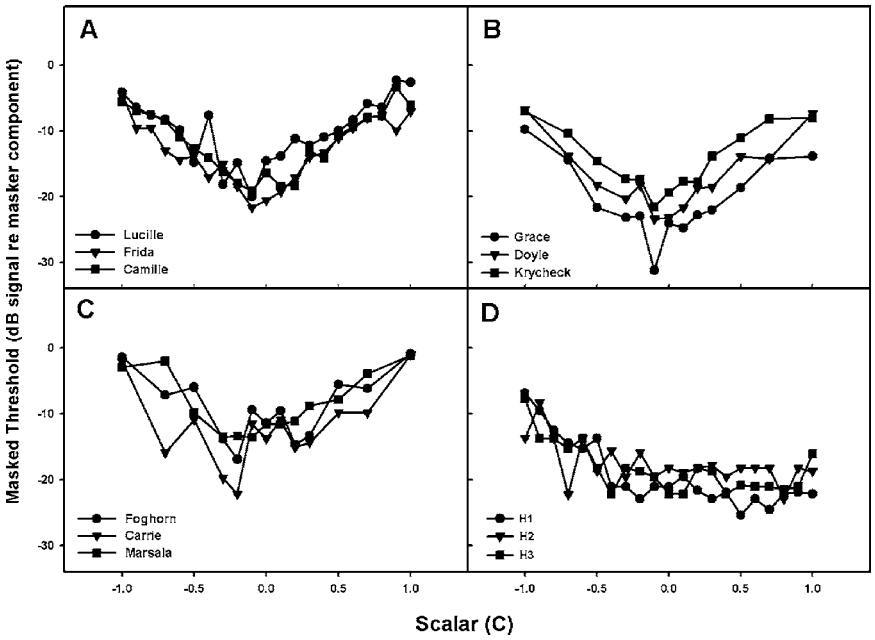


FIG. 2. Thresholds for a 2.8 kHz tone (dB signal re. masker component) embedded in harmonic complex maskers for individual zebra finches (a), budgerigars (b), canaries (c), and humans (d). The abscissa indicates the scalar value used to construct the masker waveform. Values closer to 0.0 correspond to more modulated masker waveforms; values closer to ±1.0 correspond to flatter waveforms.

from sessions with false alarm rates higher than 18% were discarded. In all, 0.7% of all sessions for the budgerigars, 5.3% for zebra finches, and 0% for canaries were excluded from analysis.

For each Schroeder-phase scalar masker, signal levels in 0.4, 1, or 2 dB steps were presented using the Method of Constant Stimuli (Dooling and Okanoya 1995). Signal levels within a condition were selected to bracket the presumed threshold, and psychometric functions were developed. Birds ran a minimum of 300 trials for each Schroeder-phase masker, and the last 200 trials once behavior stabilized (threshold did not change more than 1/3 the step size) were used for analysis. Thresholds were defined as the level of the tone detected 50% of the time, adjusted by the false alarm rate $[Pc^{*}=(Pc-FA)/(1-FA)]$ (Dooling and Okanoya 1995; Gescheider 1985). For comparison, three humans were tested with earphones on the same sounds using similar procedures. In order to estimate a phase response across frequencies, one bird of each species was also tested with signal frequencies of 1.0, 2.0, and 4.0 kHz using several of the maskers.

III. RESULTS

Figure 2 shows individual masked thresholds for a 2.8 kHz signal in dB (re. the level of each masker component) for zebra finches, budgerigars, canaries, and humans [panels (a) through (d), respectively]. Thresholds for each subject are plotted as a function of scalar value *C*. Each bird species shows a general pattern of high thresholds at the scalar extremes (+ or −1.0), with a drop in threshold near the center of the scalar range. The variability of the scalar resulting in minimum masking within bird species is quite small, and across bird species there is a systematic effect of waveform shape produced by different selections of component phase. The pattern shows a release from masking for each bird species for scalars that are just slightly negative (i.e., −0.1 and −0.2). This scalar does not produce the most highly modulated external waveform. Rather, the most highly modulated waveform is produced by a scalar of 0.0, a cosine-phase wave.

Humans show a different pattern of thresholds across scalar values than do birds. Although there are differences due to phase selection, the minimum masking for humans occurs for maskers with positive scalars. Moreover, the least amounts of masking in humans occur over several positive scalars, resulting in a much broader minimum, on average, than was observed in birds. These results are consistent with the results of an earlier study of humans by Lentz and Leek (2001) who found a minimum masking scalar for humans at a signal frequency of 3 kHz (near the signal frequency tested here) ranging between +0.5 and +1.0, but with considerable variability across subjects.

Figure 3 shows the mean values for each species. Shaded areas indicate the minimum amounts of masking.

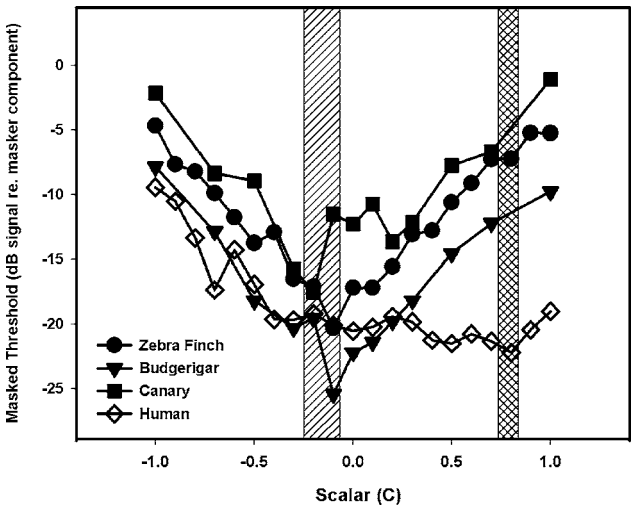


FIG. 3. Average thresholds for a 2.8 kHz tone embedded in harmonic complex maskers for zebra finches, budgerigars, canaries, and humans. The abscissa indicates the scalar value used to construct the masker waveform.

These average functions highlight the species differences in the shape of the masking functions with more sharply defined minimum masking regions for the birds and a relatively shallower and broader minimum for humans. All three species of birds show a similar pattern of masking across scalars and this pattern is different from that observed in humans. Budgerigars have overall levels of masking that match most closely those of human listeners at the negative scalar values but not at extreme positive scalar values. Zebra finches and canaries show patterns that are quite different from humans, and both show similar masked thresholds at negative and positive scalar values. A two-way repeated measures analysis of variance (ANOVA) indicated that there was a significant between-subjects effect of species $[F(3,8)=13.67, p=0.002]$ and a significant within-subjects effect of scalar $[F(12,96)=36.20, p<0.0001]$. Furthermore, there was a significant interaction of species and scalar $[F(36,96)=3.83, p<0.0001]$. A post-hoc Bonferroni t test showed that data from zebra finches and canaries were significantly different from humans ($p=0.025$ and $p=0.003$, respectively), and that budgerigar data were different from canary thresholds ($p=0.010$). In general, the release from masking across scalar values (i.e., the largest difference in threshold across thresholds) is greater in the bird species than in humans, with zebra finches showing the largest release from masking and humans the least.

As a check on the generalizability of the results obtained at a signal frequency of 2.8 kHz, we also tested one bird of each species at three other signal frequencies (i.e., 1.0, 2.0, and 4.0 kHz). Figure 4 shows that while the overall amount of masking at each signal frequency varied somewhat, all three birds still showed similar patterns of masking. The least masking occurred at 1.0 or 2.0 kHz for all species, and the most masking occurred at 4.0 kHz. These overall masking differences probably reflect the critical ratios at these frequencies, with increasing critical bandwidths at the higher frequencies for all these bird species (Okanoya and Dooling 1987).

The release from masking (maximum-minimum amount of masking) that occurs with changes in scalar value is summarized in Fig. 5 for each frequency. All three species showed the smallest release from masking due to a temporal waveform shape for a 1.0 kHz signal. The zebra finch and canary showed the largest release from masking at 4.0 kHz, while the budgerigar showed the largest release from masking at 2.8 kHz. This species difference in masking parallels species differences in other masking phenomena such as critical ratios which are larger for zebra finches and canaries than for budgerigars (Okanoya and Dooling 1987).

Interestingly, the minimum amount of masking occurs at a negative scalar for all birds and does not change considerably across frequencies in birds. These results are plotted in Fig. 6. The relatively small change in the scalar producing the least masking in birds is in contrast to results reported for humans. The data from humans taken from Lentz and Leek (2001) are plotted for comparison and show an inverse relation between the scalar value resulting in minimum masking and signal frequency. Lentz and Leek suggested that this relationship reflected the curvature in the phase-by-frequency map of the basilar membrane. The similarity in minimum masking scalar at different frequencies of the birds would argue that their phase curvatures are constant across frequency.

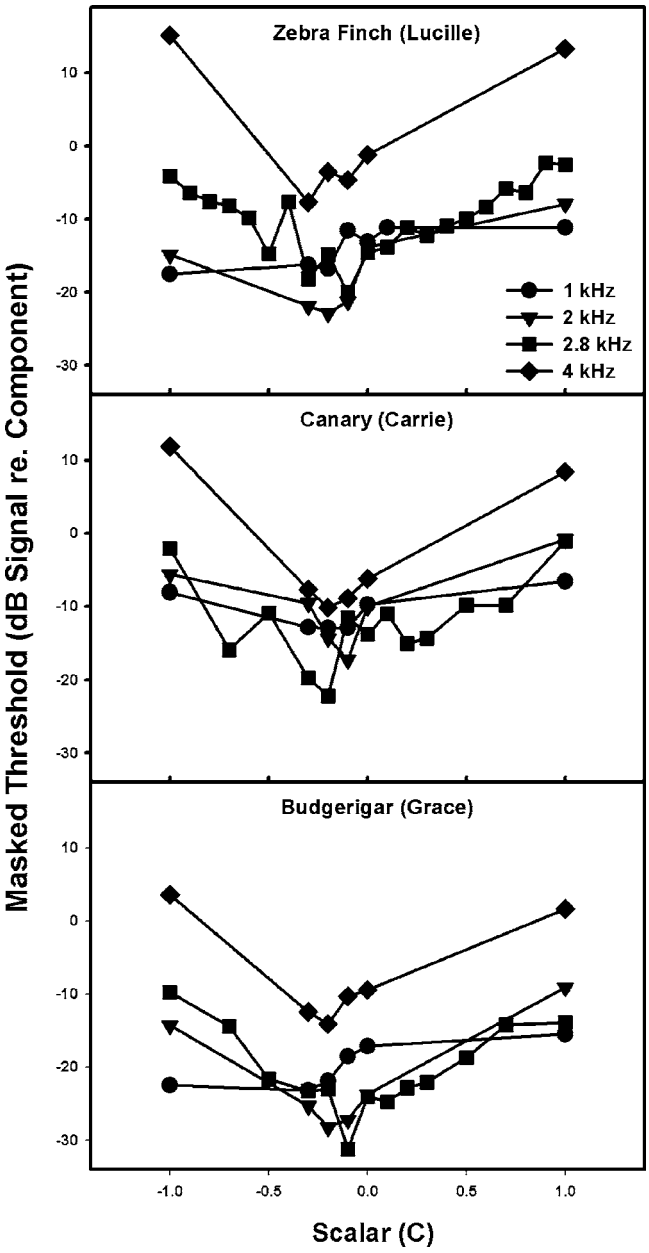


FIG. 4. Masked thresholds for 1.0, 2.0, and 4.0 kHz tones (dB signal re masker component) for a zebra finch, a canary, and a budgerigar. The abscissa indicates the scalar value used to construct the masker waveform.

IV. DISCUSSION

Thresholds were measured in three different species of small birds, as well as in humans, for tones embedded in maskers that varied in waveform shapes from highly peaked to quite flat. Thresholds in all species varied as a function of temporal waveform shape, but different patterns of masking emerged between birds and humans. The results of this study provide further evidence that temporal waveform shape affects masking in birds differently than in humans.

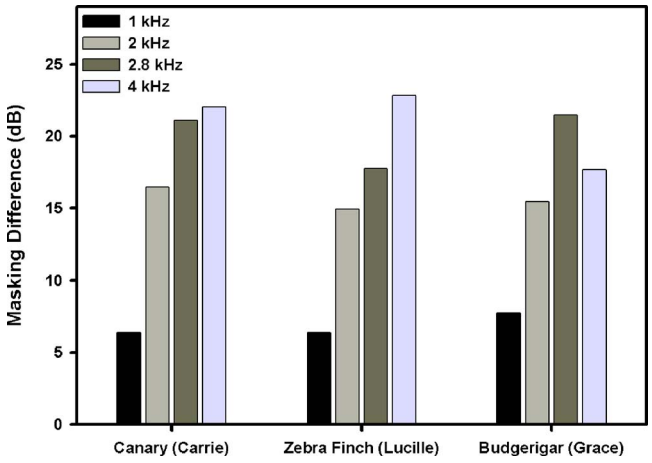


FIG. 5. Average differences in masking effectiveness of the harmonic complexes that produced the maximum and minimum amounts of masking of 1.0, 2.0, 2.8, and 4.0 kHz tones in a zebra finch, a canary, and a budgerigar.

Earlier comparative studies of Schroeder-phase masking in humans and birds have shown that in humans, differences between the original positive and negative Schroeder-phase maskers (i.e., with a scalar of ±1.0) were on the order of 15 to 20 dB, but birds' thresholds were not more than 3–8 dB apart, depending on the fundamental frequency of the masker waveforms (Leek *et al.* 2000; Dooling *et al.* 2001). Further, the more effective masker in humans was always the negative Schroeder-phase masker, but the positive-phase masker usually produced slightly more masking in birds. These two characteristics, larger differences in Schroeder-phase masking in humans than in birds and the opposite sign of the more effective masker across species, were taken to reflect basic differences in cochlear structure and function between humans and birds. Leek *et al.* (2000), however, also showed that there were some large differences in masking effectiveness in birds for harmonic complexes constructed in cosine phase (here, a scalar of 0.0) and random phase, differences

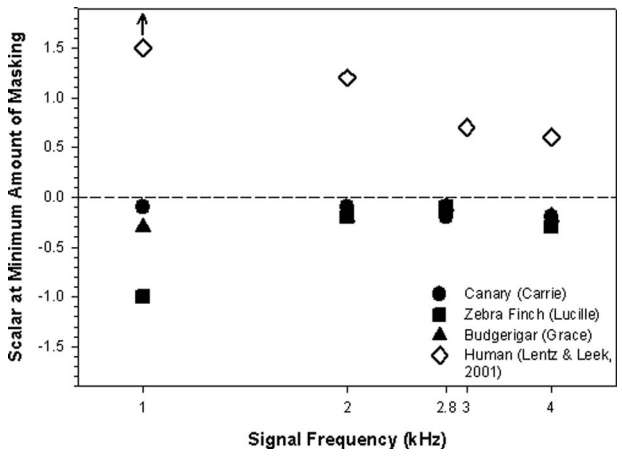


FIG. 6. Scalar values corresponding to the harmonic complexes that produced the minimum amount of masking for different frequencies in a zebra finch, a canary, and a budgerigar are compared to values reported in humans by Lentz and Leek (2001). The arrow above the data point for humans at 1.0 kHz indicates that the true masker producing the minimum amount of masking was probably greater than those reported.

on the order of 15–20 dB. Recall that random-phase waveforms are likely to have relatively flat envelopes, and cosine-phase waveforms are highly peaked.

The results from that earlier study (i.e., Leek *et al.* 2000) indicated that, under some circumstances, the waveform shape could have a large effect on masking in birds, notwithstanding the similar amounts of masking for the positive- and negative Schroeder-phase waves. This actually foreshadowed the finding here, that, given the appropriate scaled Schroeder waveform, reflecting perhaps a "matched" phase curvature in the cochlea of birds, the large Schroeder-phase differences in masking found in humans would be observed in birds. In other words, the lack of a large difference in birds between masking by positive and negative Schroeder-phase waveforms observed in earlier studies may have been because the phase spectra were incorrectly chosen for the avian cochlea. When an appropriate choice is made of monotonic phase change across frequency, and, as a result, an appropriate within-period frequency sweep rate, differences in masking effectiveness may be as large for all three bird species as in humans. This, in turn, suggests that the same mechanisms underlying these large masking differences due to waveform shape may be found in both mammalian and avian auditory processing.

In humans, the release from masking for the positive-phase Schroeder waveforms is thought to result from a cancellation between the phase of the stimulus and the phase of the auditory filter at the frequency place of the signal. This, in turn, is thought to occur because of characteristics of the traveling wave on the basilar membrane. The data reported here for birds also show masking differences in response to systematic phase changes in these harmonic complexes. Notwithstanding large differences in morphology and physiology between avian and mammalian cochleas, a similar explanation might be advanced in birds. Major differences in cochlear anatomy and mechanics include the much shorter length of the papilla, the distribution and types of hair cells, the configuration of hair cells into a matrix of support cells, but without the pillar cells found in mammalian cochleas, and the relatively larger and thicker tectorial membrane in avian ears than in mammalian ears, to name but a few [see Gleich and Manley (2000) for a comprehensive review of bird ear anatomy and physiology]. These differences are all likely to have an impact on the traveling wave and sound processing mechanisms that have been implicated in masking by harmonic complexes in humans.

Harmonic complexes with flatter temporal waveform shapes are generally more effective maskers than complexes with more peaked waveform shapes in birds and in humans. However, none of the species tested demonstrated the lowest thresholds with the peakiest masker waveform ($C=0$), but at less peaked harmonic complexes, in the positive Schroeder-phase maskers for humans and for negative Schroeder-phase complexes in all bird species. In a pattern that is consistent with previous studies that involved the original Schroeder complexes [i.e., $C=\pm 1$; (Dooling *et al.* 2001; Leek *et al.* 2000)], similar amounts of masking were observed here for the positive and negative stimuli in birds. For nearly all scaled stimuli used here, the symmetry between waveforms

scaled by plus and minus one was notable in all birds. The pattern of results for human listeners was markedly different. As in previous reports, negative- and positive-phase complexes produced very different amounts of masking, and the lowest threshold occurred at a positive-phase scalar in humans (Lentz and Leek 2001; Oxenham and Dau 2001).

Kohlrausch and Sander (1995) and Lentz and Leek (2001) have argued that the patterns of masking by Schroeder-phase complexes may be used to estimate the phase characteristic of the auditory filters centered in regions of the signal frequencies tested. For these scaled versions of the Schroeder complexes, the scalar that produces the least amount of masking at a given signal frequency indirectly indicates the rate of change of phase as a function of the rate of change of frequency across the auditory filter. That is, the second derivative of the phase-by-frequency function in the masker waveform producing the least amount of masking among these stimuli is the phase curvature in the signal frequency region of the basilar membrane (but of opposite sign). As pointed out by Kohlrausch and Sander (1995), the Schroeder-phase waveform has a constant curvature given by the quantity κ

$$\kappa = \frac{d^2\phi}{df^2} = C\frac{2\pi}{Nf_0^2}. \tag{2}$$

Applying this equation for $C=-0.1$, the scalar producing the minimum masking for budgerigars and finches, and -0.2 for canaries, gives an estimate of the cochlear phase curvature of 1.28×10^{-6} and 2.56×10^{-6}, respectively. Note that the sign of the cochlear phase curvature is opposite to the sign of the phase curvature in the waveform. For humans, the scalar value that provides minimum masking for a signal of 3.0 kHz is +0.75 with the same masker stimuli used in this study (Lentz and Leek 2001), so the curvature of the cochlear phase characteristic in that frequency region is estimated to be -9.62×10^{-6}. These comparative data suggest that the phase change across frequency in humans is at least four times as rapid as in birds and is in the opposite direction.

The differences in the curvature of the phase characteristic are also reflected in the temporal features of the traveling wave in birds and humans. The species differences in masking patterns observed in this study might be explained by examining the interaction between those features and the phase spectra of the stimuli. Based on the cochlear frequency map provided by Gleich (2000) for the starling, the signal frequency of 2.8 kHz used here would fall at about 0.8 mm from the base of the cochlea. According to traveling wave velocities reported by (Gleich 2000), the time for a stimulus to arrive at the appropriate region of the cochlea would be about 0.4–0.5 ms. In humans, in contrast, the place on the cochlear map that processes 2.8 kHz is about 14 mm from the base (Greenwood 1961), and the time required for the traveling wave to arrive there is some 6.5 to 7.5 ms (Donaldson and Ruth 1993).

The minimum masking produced by a Schroeder-phase complex should occur in the region where the movement of the cochlear partition maximally compensates for the waveform shape of the external stimulus to produce a highly modulated internal masker. This should occur for a scaled Schroeder-phase masker that matches the time of arrival of the signal component in the stimulus with the time of arrival of the traveling wave in the cochlea, but with a reversed sign to just compensate for the traveling wave. Thus, in birds we need a scaled Schroeder masker in which the 2.8 kHz component arrives at about 0.5 ms into the period of the waveform, and for humans, we require a masker in which the 2.8 kHz component arrives about 7 ms into the period. Recall the stimulus characteristics: frequencies ranging from 200–5000 Hz, with a fundamental period of 10 ms. Recall also that as the absolute values of the scalars approach 0.0, the rate of change of frequency increases. These facts, along with the recognition that the signal frequency of 2.8 kHz is clearly about midway into the frequency range of the masker, indicate that the minimum masker for birds will be a much smaller scalar than that found for humans. In fact, by observing the time course of the frequencies within each period of the masker for each scalar, we may calculate that a scalar of -0.1 will produce a masker that reaches the 2.8 kHz component in about half a millisecond, and, perhaps significantly, it is that scalar that provides the least amount of masking in most of the birds in this study. A scalar of +0.5 to +0.7 results in maskers in which the 2.8 kHz component arrives from 6–7 ms into the period, and it is this region that provides the least amount of masking in humans. Thus, to a first approximation, the time of arrival of the signal frequency in the masker might be thought to just compensate for the time of arrival of the traveling wave in response to the masker-plus signal. This would suggest that a harmonic complex with frequencies arranged to occur in a glide lasting no more than the maximum latency from base to apex of a given cochlea would provide the best estimate of travel time in the cochlea. The complexes used here were 10 ms in glide time (within period). Perhaps something more on the order of 1–3 ms would better suit the shorter cochleas of birds, and in fact the glide times in the scaled stimuli with a scalar of 0.1 to 0.3 requires approximately that much time within each period to extend across the frequency range, regardless of the actual period of the stimulus. In other words, perhaps the bird papillae are simply too short to support different phase alterations across frequency. The papilla of zebra finches and canaries is 1.6 mm, and it is about 2.1 mm for budgerigars (Gleich *et al.* 1994; Manley *et al.* 1993). Thus, an upward-sweeping within-period glide occurring in about 1.03 ms in zebra finches and canaries and about 1.3 ms in budgerigars may be the temporal limits for introducing a phase change, compared to around 10 ms in humans.

Gleich and Manley (2000) proposed that the main stimulus to the hair cells in the avian basilar papilla is likely related to the resonance of the tectorial membrane (TM), and is indirectly influenced by the vibratory motion of the basilar membrane. The frequency-dependent motion of the basilar membrane would first activate those cells over it, the efferently innervated hair cells, producing active movement in the stereociliary bundles. This movement would feed mechanical energy into the TM in phase with the stimulus. Movement of the TM would then activate afferently innervated hair cells. According to this model, altered representations in response

to different input phases must occur subsequent to the motion of the basilar membrane. The near-zero scalars that produce minimum masking in the birds for all three frequencies tested might suggest that there is little phase change imposed by cochlear processing in the bird, which would be consistent with only small changes to basilar membrane motion, in contrast to the large and frequency-dependent phase lags observed in mammalian cochlear processing.

Taken together, the comparative results on masking by harmonic complexes by birds and humans along with previous findings of enhanced discrimination of temporal fine structure in harmonic complexes by birds (Dooling *et al.* 2002) invites speculation about a match between the temporal features of bird song and auditory specializations for perceiving song. The vocalizations of many birds are well characterized by temporal precision, rapid frequency sweeps, and in some cases, complex harmonic patterns (Greenewalt 1968). There is strong evidence for species-specific perceptual specializations enhancing the perception of species-specific calls (Dooling and Searcy 1979; Okanoya and Dooling 1991). More recent investigations show that birds are acutely sensitive to changes in the temporal fine structure of these natural vocalizations (Lohr *et al.* 2000) and evidence from studies of the neuromuscular activity of the syrinx show control of extremely fine temporal and spectral detail (see Suthers and Zollinger 2004 for review). Thus, the curious differences between humans and birds in the processing of these harmonic complexes may be a reflection of auditory processing especially suited for decoding the fine detail in bird vocalizations.

V. CONCLUSIONS

These data provide a psychophysical description of the phase response of the avian basilar papilla. As in previous reports of masking by Schroeder-phase harmonic complexes in birds, we have shown that temporal waveform shape can affect masking in birds and humans in very different ways. Further, we have demonstrated that the least effective scalar Schroeder-phase masker in birds is different from that of humans at the same frequency channel. In birds, the least effective masker has a scalar value close to zero and has a negative phase. In humans, the least effective masker varies somewhat, but is always a positive phase. These results likely reflect fundamental anatomical and physiological differences in cochlear phase response between birds and humans, in particular differences in lengths of the basilar membrane and basilar papillae, and differences in delay as a function of frequency of the traveling wave. These differences may underlie some demonstrated cases of birds being able to hear fine detail in their vocalizations to which humans are insensitive.

ACKNOWLEDGMENTS

This work was supported by NIH Grants DC-01372 to R.J.D., DC-005450 to A.M.L., DC-00626 to M.R.L., and DC-04664-01A2 (Core Center grant) to the University of Maryland. Special thanks to Dr. Micheal Dent for help on various aspects of these studies. 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 the Department of Defense. The authors thank Otto Gleich for insightful discussions and comments about this research.

Carlyon, R. P., and Datta, A. J. (**1997a**). "Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system," J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Datta, A. J. (**1997b**). "Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components," J. Acoust. Soc. Am. **101**, 3648–3657.

Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.

Dooling, R. J., Dent, M. L., Leek, M. R., and Gleich, O. (**2001**). "Masking by harmonic complexes in birds: Behavioral thresholds and cochlear responses," Hear. Res. **152**, 159–172.

Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (**2002**) "Auditory temporal resolution in birds: Discrimination of harmonic complexes," J. Acoust. Soc. Am. **112**, 748–759.

Dooling, R. J., and Okanoya, K. (**1995**). "The Method of Constant Stimuli in testing auditory sensitivity in small birds," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhauser-Verlag, Basel), pp. 161–169.

Dooling, R. J., and Searcy, M. (**1979**). "Early perceptual selectivity in the swamp sparrow," Dev. Psychobiol. **13**, 499–506.

Gescheider, G. A. (**1985**). *Psychophysics: Method, Theory, and Application* (Lawrence Erlbaum and Associates, New York).

Gleich, O. (**2000**). Group delay and the velocity of excitation along the auditory sensory epithelium of starling and guinea pig," Assoc. for Res. in Otolaryngol. (ARO, St. Petersburg Beach, FL).

Gleich, O., and Manley, G. A. (**2000**). "The hearing organ of birds and crocodilia," in *Comparative Hearing: Birds and Reptiles* edited by R. J. Dooling, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 70–138.

Gleich, O., and Narins, P. M. (**1988**). "The phase response of the primary auditory afferents in a songbird (*Sturnus vulgaris*)," Hear. Res. **32**, 81–92.

Gleich, O., Manley, G. A., Mandl, A., and Dooling, R. J. (**1994**). "Basilar papilla of the canary and zebra finch: A quantitative scanning electron microscopical description," J. Morphol. **221**, 1–24.

Greenewalt, C. H. (**1968**). *Bird Song: Acoustics and Physiology* (Smithsonian Inst. Press: Washington, DC.

Greenwood, D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

Kohlrausch, A., and Sander, A. (**1995**). "Phase effects in masking related to dispersion in the inner ear II. Masking period patterns of short targets," J. Acoust. Soc. Am. **97**, 1817–1829.

Leek, M. R., Dent, M. L., and Dooling, R. J. (**2000**). "Masking by harmonic complexes in budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **107**, 1737–1744.

Lentz, J. J., and Leek, M. R. (**2001**). "Psychophysical estimates of cochlear phase response: Masking by harmonic complexes," J. Assoc. Res. Otolaryngol. **2**, 408–422.

Lohr, B., Bartone, S., and Dooling, R. J. (**2000**). "The discrimination of fine-scale temporal changes in call-like harmonic stimuli by birds," *Association for Research in Otolaryngology* (ARO, St. Petersburg, FL).

Manley, G. A., Scwabedisen, G., and Gleich, O. (**1993**). "Morphology of the basilar papilla of the budgerigar (*Melopsittacus undulatus*)," J. Morphol. **218**, 153–165.

Okanoya, K. and Dooling, R. J. (**1987**). "Hearing in passerine and psittacine birds: A comparative study of absolute and masked auditory thresholds," J. Comp. Psychol. **101**, 7–15.

Okanoya, K., and Dooling, R. J. (**1991**). "Perception of distance calls by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*): Assessing species-specific advantages," J. Comp. Psychol. **105**, 60–72.

Oxenham, A. J., and Dau, T. (**2001**). "Towards a measure of auditory-filter phase response," J. Acoust. Soc. Am. **110**, 3169–3178.

Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.

Schroeder, M. R. (**1970**). "Synthesis of low-peak-factor signals and binary

sequences with low autocorrelation," IEEE Trans. Inf. Theory **16**, 85–89.
Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). "Phase effects in masking related to dispersion in the inner ear," J. Acoust. Soc. Am. **80**, 1631–1637.
Summers, V., and Leek, M. R. (**1998**). "Masking of tones and speech by Schroeder-phase harmonic complexes in normally-hearing and hearing-impaired listeners," Hear. Res. **118**, 139–150.
Suthers, R. A., and Zollinger, S. A. (**2004**). "Producing song: the vocal apparatus," Ann. N.Y. Acad. Sci. **1016**, 109–129.
von Bekesy, G. (**1960**). *Experiments in Hearing* (McGraw Hill, New York).

Preliminary evidence for the use of microseismic cues for navigation by the Namib golden mole

Edwin R. Lewis[a)]
Department of Electrical Engineering and Computer Science, University of California, Berkeley, California 94720-1771

Peter M. Narins
Department of Physiological Science, University of California, Los Angeles, California 90095-1606

Jennifer U. M. Jarvis and Gary Bronner
Department of Zoology, University of Cape Town, Cape Town, South Africa

Matthew J. Mason
Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom

(Received 29 August 2005; revised 14 November 2005; accepted 18 November 2005)

Insect prey of the Namib golden mole congregate beneath clumps of grass scattered among the sand dunes of the Namib Desert. In the presence of the light winds that typically blow over the Namib Desert, these grass clumps emit low-amplitude vibrations that are transmitted through the sand. While foraging in the sand-swimming mode (a few centimeters below the surface of the sand), some moles apparently were attracted toward manmade sources emitting vibrations matching those recorded from the grass clumps. This is the first direct evidence that these desert mammals use seismic cues for navigation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2151790]

PACS number(s): 43.80.Lb, 43.66.Gf, 43.64.Tk [JAS] Pages: 1260–1268

I. INTRODUCTION

There is growing literature on the uses or potential uses of seismic signals (substrate vibrations), especially by arthropods (Autrum and Schneider, 1948; Markl, 1983; Cocroft, 2001) and fossorial mammals (Mason and Narins, 2001; Kimchi and Terkel, 2002), but also by reptiles (Hartline, 1969, 1971; Hetherington, 1989, 1992; Young and Morain, 2002), amphibians (Koyama *et al.*, 1982; Lewis, 1984; Lewis and Narins, 1985; Lewis *et al.*, 2001), and even large land mammals (O'Connell-Rodwell, Hart, and Arnason, 2001). Use of seismic cues for prey localization in sandy habitats has been a productive topic (Brownell, 1977; Brownell and Farley, 1979a, 1979b, 1979c; Hetherington, 1989, 1992; Young and Morain, 2002), and one especially relevant to the work described here.

Among the dunes of the Namib Desert, clumps of dune grass (*Stipagrostis sp.*) or ostrich grass (*Cladoraphis spinosa*) form small, scattered islands on a sea of sand. Beneath these islands, living among the roots of the grass, are dune termites (*Psammotermes sp.*), the principal insect prey of the Namib golden mole (*Eremitalpa granti namibensis*) (Fielden, Perrin, and Hickman, 1990). While foraging for these insects, the mole, which has poorly-developed eyes with fused eyelids (Gubbay, 1956) and cannot form images, takes remarkably straight paths from grass clump to grass clump. These paths frequently span more than ten meters between clumps, and they cover an average distance of approximately 1400 m per night (Fielden *et al.*, 1990; Fielden, 1991; Seymour, Withers, and Weathers, 1998). A previous study, carried out in 1993 near Gobabeb, demonstrated that this foraging is purposeful, not random (Narins *et al.*, 1997). Foraging must, therefore, involve sensory cues (other than vision), even if the animal learns its foraging routes and then reuses them.

Certain golden moles, including *Eremitalpa*, are well-known to have disproportionately large auditory ossicles (Forster Cooper, 1928; von Mayer, O'Brien, and Sarmiento 1995; Mason, 2003b). The combined mass of the malleus and incus in *Eremitalpa* is approximately 0.1% of the animals total body mass, compared to only 0.001% in the comparably-sized laboratory mouse, *Mus musculus*, and only 0.00008% in humans (see Mason, 2001). Several authors have suggested that these hypertrophied ossicles are adaptations for detecting ground vibrations (Kuyper, 1984; Fielden *et al.*, 1990; Hickman, 1990; Narins *et al.*, 1997), perhaps by means of a mechanism referred to as *inertial bone conduction* (Fay and Popper, 1985; Lombard and Hetherington, 1993). Employing a physical model of inertial bone conduction, Mason investigated this possibility further (Mason and Narins, 2002; Mason, 2003a). He concluded that the geometries of the enlarged ossicles of *Eremitalpa* made them especially suitable for the detection of low-frequency ground-borne vibrations of very low amplitude (low-frequency microseismic signals). The golden mole might use this ability in order to navigate, and possibly also to identify and locate prey on or beneath the sand (Narins *et al.*, 1997).

In its natural habitat, E.g., *namibensis* moves either on the surface of the sand or just beneath it with a sand-swimming motion (Meester, 1964; Holm, 1969; Fielden,

[a)]Electronic mail: lewis@eecs.berkeley.edu

Hickman, and Perrin, 1992). When it is on the surface, the mole stops approximately every 60 cm and dips its head into the sand. As it comes close to a grass clump, its head dips occur closer together (Narins *et al.*, 1997). Head-dipping by desert vertebrates is believed to couple the head more closely to the substrate, aiding in the detection of microseismic signals (Hetherington, 1992; Narins *et al.*, 1997). In the sand-swimming mode of travel, beneath the surface of the sand, the mole's head would be coupled continuously to the substrate. The earlier study at Gobabeb demonstrated that the mole's preferred microhabitats for feeding (clumps of live dune grass) emit distinct microseismic signals whenever there is wind (Narins *et al.*, 1997). Moles typically rest beneath dune-grass clumps when they are not feeding, thus using them for shelter as well. We hypothesized that if E.g., *namibensis* does use microseismic cues for navigation, then it should be attracted to buried artificial seismic sources emitting appropriate signals. We tested this hypothesis with moles released in the presence of seismic transducers emitting the vibration patterns of dune grass clumps in a gentle wind (thus simulating familiar landmarks).

II. METHODS

We carried out our study on Die Duine Farmstead (25° 14.47′ S, 16° 03.39′ E), approximately 100 km west of Maltahöhe, Namibia, in the NamibRand Nature Reserve. Approximately 1 km from the farmstead buildings, we selected a flat area of sand, clear of grass clumps and approximately 20 m in diameter, at the center of which we planted a small clump of ostrich grass (*Cladoraphis spinosa*). We established an experimental arena as the area inside a circle, 10 m in diameter and centered on the clump of ostrich grass. Around the perimeter of the arena we deployed eight seismic sources at 45° intervals, beginning at a position 22.5° east of north (i.e., north-northeast) relative to the center of the arena. These sources (Clarke Synthesis transducers, model TST 229 F4 ABS) were disk-shaped, 21 cm in diameter and were buried approximately 30 cm below the surface of the sand. For each source, a uniquely labeled coaxial cable was laid outside the arena. The cables were buried approximately 10 cm beneath the surface of the sand, and extended to a small table approximately 5 m beyond the edge of the arena and west-northwest from the arena's center. A vertical geophone (OYO Geo Space GS-20DX, damped with a 1-kΩ resistor) was buried to a mean depth of approximately 4 cm near the center of the arena, on the side of the ostrich-grass clump closest to the table. Its coaxial cable was buried approximately 15 cm below the surface of the sand and led directly through the arena to the table. The geophone cable was connected to a high-gain, low-noise amplifier (constructed from National Semiconductor LF441 ACH operational amplifiers), the output of which was connected to one input channel of a battery-operated oscilloscope (Sony/Tektronix 305).

The signal for driving the seismic sources was synthesized digitally and recorded as a 16-bit wave file on a compact disk (CD). It comprised 60 s of bandlimited Gaussian noise, the amplitude spectrum of which had been sculpted to match the difference spectrum from the earlier Gobabeb recordings [Narins *et al.*, 1997, Fig. 5(b)]. This was the difference between the amplitude spectrum of the seismic noise recorded (in a gentle wind) from the sand immediately adjacent to dune-grass clumps and that of the seismic noise recorded simultaneously in open areas at least 20 m away. In our last two experiments, we used the same noise pattern with randomly-timed click waveforms added to it. We had recorded these clicks from a vertical geophone buried to a mean depth of approximately 4 cm at the edge of an ostrich-grass clump approximately 200 m west of the arena. Because the clicks occurred near only a small fraction of the ostrich-grass clumps that we sampled, we surmised that they were emissions from potential prey items (e.g., insects) of the mole. The stimulus CD was played from a portable CD player (Classic CL413), with a track repeat feature. Thus the 60-s track was repeated with almost seamless transitions for many hours. The selected output channel (right or left) of the CD player was connected to both input channels (right and left) of a stereo power amplifier (Road Gear RGCA 100) and to a single input channel of a second, identical amplifier. Each of the corresponding outputs was connected through one of the buried coaxial cables to a single seismic source at the perimeter of the arena.

For each experimental trial, three neighboring sources were activated; the other five sources were silent. The amplitude of each source was adjusted to yield approximately 0.0001 cm/s (root-mean-square vertical velocity) at the geophone at the center of the arena. Within the arena, the vibration field established by this excitation pattern could be divided into two regions of approximately equal areas. In one region the stimulus vibration amplitude (from the active sources) was greater than that at the point at which each mole was released; in the other region it was less than that at the release point. The trio of sources to be made active was varied randomly from trial to trial. At the adjusted vibrational level, the air-borne sound level, measured 0.5 m from the surface of the sand directly above each source was approximately 36 dB SPL. The vibrational stimulus spectrum was centered at approximately 350 Hz [see Fig. 5(b) in Narins *et al.*, 1997]. Evidently owing to frequency-dependent coupling from sand to air, the spectrum of the corresponding airborne sound was centered at a lower frequency, approximately 240 Hz. We recorded airborne sound emitted by a *Stipagrostis* clump in a gentle breeze and found that its power was concentrated between 35 and 90 Hz, with much reduced levels above 100 Hz.

For each trial, once the source amplitudes had been adjusted, a single mole was released at approximately the same position (the west-southwest edge of the ostrich-grass clump at the center of the arena). Upon release, the mole immediately burrowed into the sand near or beneath the clump and remained there for several hours. The investigators then withdrew as quietly as possible, to the farmstead buildings in most cases. In those cases, the mole's activity was checked periodically by a single observer approaching the arena stealthily. In a few cases, one or two investigators remained behind, sitting very quietly at a distance of more than 50 m from the arena. From there, the mole's activity was observed through binoculars. In every trial, the path that the mole had

TABLE I. Basic information is given about each of the eleven moles, numbered in order of release. Neither the sex nor the weight of mole 1 was determined before its release. GMT is Greenwich Mean Time.

Mole number	Date caught	Sex	Mass (g)	Date, local time (GMT+3) of release
1	Sep 01, 2002	···	···	Sep 01, 1703 h
2	Aug 30, 2002	M	20	Sep 02, 1158 h
3	Aug 31, 2002	M	25	Sep 02, 1920 h
4	Aug 31, 2002	F	18	Sep 03, 1235 h
5	Aug 31, 2002	M	22	Sep 03, 2258 h
6	Aug 31, 2002	M	26	Sep 06, 1130 h
7	Aug 31, 2002	F	15.5	Sep 07, 1210 h
8	Sep 07, 2002	F	19	Sep 08, 1043 h
9	Sep 07, 2002	F	16.5	Sep 09, 1347 h
10	Sep 07, 2002	F	19	Sep 12, 1530 h
11	Sep 08, 2002	M	19.5	Sep 13, 1610 h

taken was easily seen. Once the mole had exited the arena and left the vicinity, its track was mapped in a piecewise-linear fashion. The heading of each track segment was measured (with a string or a narrow-beam light) relative to the arena perimeter, and the length of each segment was measured. Preparation for the subsequent trial included thorough raking of the sand, down to approximately 10 cm, to erase the previous mole's tracks and to disperse any scent-trail it may have left.

Eleven moles were available for this project (see Table I); each was used only once and subsequently returned to the wild. While in captivity, the moles were housed in 20-liter plastic buckets containing dune sand and were provided with mealworms (tenebrionid larvae) and local termites ad lib.

III. RESULTS

The results for all eleven moles are presented in Figs. 1–4, which show the mole tracks superimposed on the vibration field computed (on the basis of two-dimensional power spread) for the corresponding trio of active sources. The right-hand panel of Fig. 4 shows the orientation of the arena, determined by GPS observations, in each of the other panels in Figs. 1–4. The axes are labeled in meters from the center of the arena. The large outer circle represents the perimeter of the arena; the small, inner circle represents the ostrich grass clump (approximately to scale). The sources are depicted as small circles (filled with black and drawn to scale) distributed along the large outer circle. Solid line segments (gray or black) represent track segments made by sand swimming moles; dotted line segments represent track segments made by moles walking on the surface. Taking geophone measurements at 32 points distributed uniformly in azimuth (45° apart) and radial distance from the center of the arena (1.0 m apart), we corroborated and calibrated the computed vibration field map to within 1.0 m of the active sources. The contour lines in Figs. 1–4, calibrated in this manner, are separated by 4 dB. In each figure, note the second innermost contour line surrounding all three active sources (about half way from the arena center to the centermost active source). This line corresponds to the vibration amplitude measured next to live dune-grass clumps under conditions of light wind

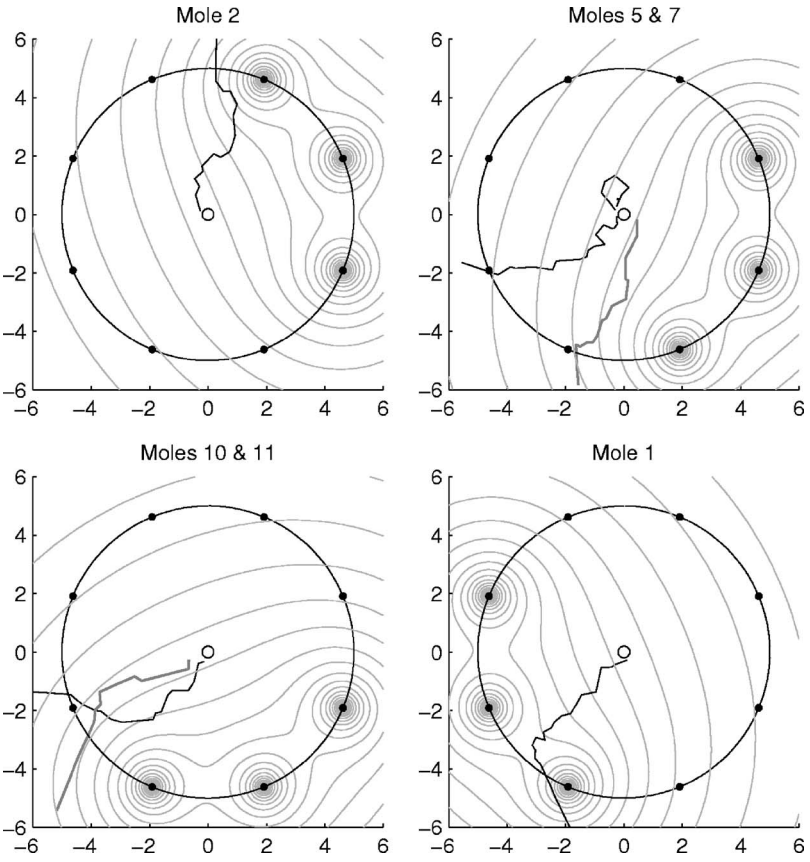


FIG. 1. Tracks of six sand-swimming moles. The axes are labeled in meters from the center of the arena. Computed contours of constant vertical-vibration amplitude are depicted in pale gray. At locations within the arena that are greater than 1.0 m from any active source, the calibrated contour spacing is 4 dB. Compass directions are given in the right-hand panel of Fig. 4. In each panel, true north is up and east is to the right. The tracks of moles 5 and 11 are depicted in dark gray, those of moles 7 and 10 in black.

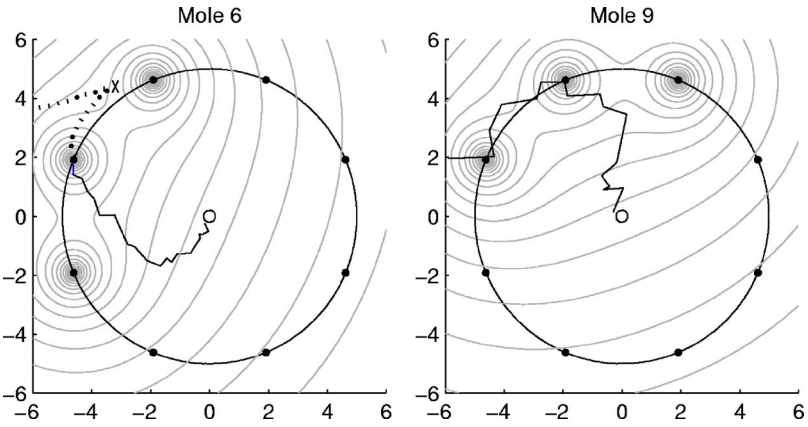


FIG. 2. Tracks of moles 6 and 9. The solid black line segments represent sand-swimming tracks. The dotted line segments represent surface-walking tracks. The final approach of mole 6 to the source in the west-northwest position was a dive, too deep to leave a track on the surface. Axes, contours, and arena orientation are the same as in Fig. 1.

(wind velocity near 0 m/s at ground level to approximately 3 m/s at 2 m above the ground) (Narins *et al.*, 1997).

For each of the eleven moles, the time between its release and its subsequent journey out of the arena ranged from 2 to 20 h. We intended to present the stimulus continuously during that time. Occasionally, however, we were required to interrupt the stimulus briefly in order replace nearly-spent batteries. Each power amplifier was powered by a pair of rechargeable, 12-V lead-acid batteries. As long as the battery voltage exceeded approximately 10 V, the amplifier performance was unchanged. When the voltage fell below that level, the amplifier shut down completely. On each visit to the arena, we checked the batteries with a voltmeter and replaced them as needed. On only one occasion (mole number 4 in Table I) did an amplifier shut down during an experiment. During a visit to the arena, the observer found that the mole had not begun to move and that the amplifiers still were functioning. On the subsequent visit, 75 min later, he found that the mole had left the arena area and that the amplifier driving the source at north-northwest and the source at north-northeast had shut down (see Fig. 4). Because we were unsure of the point at which failure had occurred, we eliminated this mole from consideration in our acoustical results. To avoid failure of the portable CD player, the dry cells powering it were replaced on a regular schedule. In only one instance (mole number 6 in Table I) did the player fail prematurely. On one visit to the arena, the observer found that the mole had moved approximately 2 m along its track toward the arena perimeter and that all three sources were functioning. On the subsequent visit, 100 min later, he found that the mole had left the arena area and that the CD player (driving all three sources) had shut down. The point at which failure occurred seemed to be indicated by an abrupt course reversal (marked by X in the left-hand panel of Fig. 2). Furthermore, the mole's conspicuous behavior in the vicinity of the west-northwest source (see below) strongly suggested that, as it approached the source, it was aware of the source's presence. Therefore, we included this mole in our statistical analysis. The reader should remain aware of the potential bias in this decision and feel free to treat the results from both moles 4 and 6 as he or she sees fit.

By the time we were ready to use moles 10 and 11 (our last two moles), we had noticed that moles tended to go directly over the active sources when they were distributed along the northwestern quadrant of the perimeter. No moles had been enticed to exit even close to active sources along the southeastern quadrant. For that reason, we tried twice more to draw the moles in the southeastern direction, adding putative prey sounds to enhance to attractiveness of the stimulus.

Most of the track plots are self-explanatory. The plots for moles 6 and 9 warrant elaboration, however. The track of mole 6 not only exited the arena at an active source (the west-northwest source), but in the final approach to that source, this mole dove toward the center of the source. The mole then resurfaced (to normal sand-swimming depth) and circled the perimeter of the source, which was buried at least 20 cm below the mole's body. After circling the source, it immediately emerged to walk on the surface, heading first toward a neighboring source (which was or had been active),

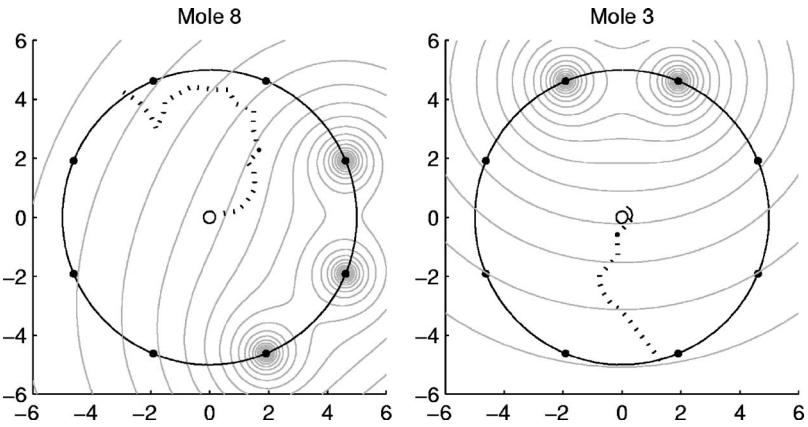


FIG. 3. Tracks of two surface-walking moles. In the case of mole 3, the third source was inadvertently disconnected from its power amplifier. Axes, contours and arena orientation are the same as in Fig. 1.

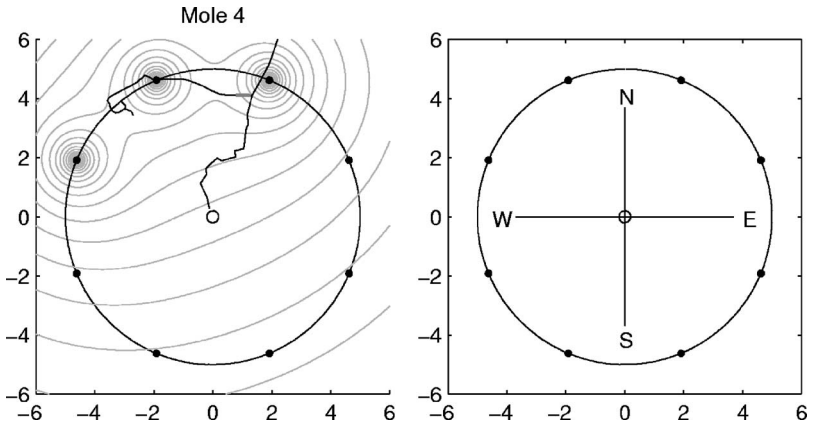


FIG. 4. (Left panel) Sand-swimming tracks of mole 4. The westward detour began with a dive (denoted by the dark gray segment), leaving no surface track. The contours show the position of the stimulus field prior to battery failure. (Right panel) Compass directions for all panels in Figs. 1–4.

then abruptly reversing direction (at the position marked by X) and heading westward, away from the arena. Mole 9 exited the arena directly over an active source, then turned westward along the arena perimeter, reentered the arena and exited again, over the neighboring active source. Then, continuing to sandswim, it headed westward, away from the arena.

IV. DISCUSSION

A. Implications regarding the use of acoustic cues

Among the ten moles counted in this study (excluding mole 4), only two (moles 6 and 9) exited the arena directly over an active source. Consider each mole to be a trial, and take its first exiting from the arena directly over an active source to be a success (a hit) and first exiting the arena anywhere else a failure (a miss). Assume, as a null hypothesis, that all points around the arena perimeter are equally likely to be the mole's point of first exit, even in the presence of the vibratory stimuli. In that case, the probability that any given trial would lead to a hit (in the presence of three active sources, each of diameter 21 cm on a perimeter with 5-m radius) is given by

$$p = (3 \times 0.21)/(2\pi \times 5) = 0.0201.$$

The probability that none of the ten trials results in a hit is given by

$$(1 - p)^{10} = 0.8166.$$

The probability that the ten trials would lead to exactly one hit is given by

$$10p(1 - p)^9 = 0.1671.$$

Therefore, the probability (P) of more than one hit in ten trials (our outcome) under this null hypothesis is

$$P = 1 - 0.8166 - 0.1671 = 0.0163.$$

This value of P gives us slightly better than 98% confidence that our outcome would not be the result of chance under the first null hypothesis. In other words, when the targets are tiny, as they are here, even two hits out of ten trials can be significant.

We cannot be sure, however, that there was no preexisting bias (independent of source position) influencing the mole's exit point from the arena. Regardless of the position (or absence, possibly, in the case of mole 4) of the active sources, none of the 11 moles exited over the south-southeast source or the north-northeast source, or over the eastern perimeter segment lying between those two sources. This was the side of the arena facing the edge of the dune fields. Extended for approximately 100 m, a path out of this part of the perimeter would carry a mole beyond the dunes, onto a vast gravel plane. The eight sources divide our arena perimeter into an ordered set of eight segments. No exit points occurred along three contiguous members of that set. Under our first null hypothesis, the probability that, after 11 trials, there remains, somewhere along the perimeter, a subset of three or more contiguous segments in which no exit points have occurred is 0.042 (see Appendix). Thus there is reason to believe that our first null hypothesis is inappropriate. A bias toward the west would increase the probability of hitting a target by chance, and thus possibly reduce the significance of the hits by moles 6 and 9. It is appropriate, therefore, to re-evaluate the chance occurrences under an alternative null hypothesis; that exiting anywhere along the perimeter segment between and including the four east-side sources has an *a priori* probability of zero, and that all points along the remaining perimeter segment are equally likely to be the mole's first point of exit.

Under this second null hypothesis, moles 2, 5, 7, and 8 had no targets at all in the acceptable region. Moles 3, 10, and 11 each had one target in the acceptable region. Mole 9 had two targets in the acceptable region, and moles 1 and 6 each had three targets in the acceptable region. In other words, the ten trials are reduced to six. The probability that any of the three trials involving moles 3, 10, or 11 would lead to a hit is given by

$$p = p_1 = 0.21/((5/8) \times 2\pi \times 5) = 0.0107.$$

For the trial involving mole 9

$$p = p_2 = (2 \times 0.21)/((5/8) \times 2\pi \times 5) = 0.0214.$$

For the two trials involving moles 1 or 6

$$p = p_3 = (3 \times 0.21)/((5/8) \times 2\pi \times 5) = 0.0321.$$

The probability that none of the six trials would produce a hit is given by

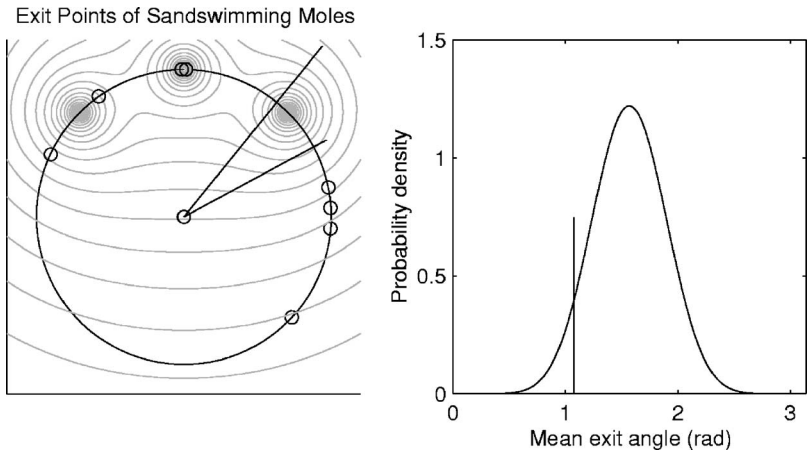


FIG. 5. (Left panel) The arena is depicted with the central active source at the top. Open circles depict the first points at which the eight sand-swimming moles exited the arena relative to the location of that source. Treating each exit point as a unit radial vector relative to an origin at the center of the arena, one can compute a vector sum (whose direction is depicted by the longer of the two radial lines). The magnitude of the vector sum is 3.51. Taking the null hypothesis to be uniform *a priori* probability distribution for each unit vector over the $-\pi$ to π rad of possible directions (θ) relative to that of the central source, one would apply the Rayleigh test. In this case, the statistic Z for the Rayleigh test has the value of 1.54. This corresponds to a P value of approximately 0.21. With eight trials, for a P value of 0.05 or less, Z would need to be 2.9 or more, corresponding to a vector-sum magnitude of 4.85 or more. The same null hypothesis can be translated to a uniform *a priori* distribution (from 0 to π) of absolute directions (abs[θ]) relative to the direction of the central source. The shorter of the two radial lines shows the direction of the mean absolute direction (mean exit angle). (Right panel) The mean exit angle (1.0765 rad) plotted against the probability density function (unit=1.0 rad^{-1}) computed for eight trials under the null hypothesis. In this case, the P value for the data is 0.063.

$$(1-p_1)^3(1-p_2)(1-p_3)^2 = 0.8877.$$

The probability that they would produce exactly one hit is given by

$$(3p_1)(1-p_1)^2(1-p_2)(1-p_3)^2 + (1-p_1)^3(p_2)(1-p_3)^2 + (1-p_1)^3(1-p_2)(2p_3)(1-p_3) = 0.1070.$$

Therefore, the probability (P) of more than one hit in the six trials (our outcome) under the second null hypothesis is

$$P = 1 - 0.8877 - 0.1073 = 0.0053.$$

This value of P gives us slightly better than 99% confidence that our outcome would not be the result of chance under the second null hypothesis. Thus the westward bias suggested by the data does not reduce the significance of the two direct hits on west-side targets.

Among the eight trials (excluding mole 4) in which the moles traveled out of the arena predominantly in the sand-swimming mode, the choice of a northerly or southerly direction appears to be correlated with the location of the trio of active sources. This was true even when the active sources were located along the (possibly unacceptable) eastern 3/8 of the arena perimeter. When the majority of active sources was located along the northern half of the perimeter, the moles (N=3) exited there; when the majority was located along the southern half, the moles (N=5) exited there. On the other hand, under the first null hypothesis (all exit points around the perimeter equally likely), a more objective assessment of directional tendency is provided by circular statistics (see Fig. 5). For both vector sum and mean angle, the data fail to reach the 95% confidence level, although the mean angle is close, at P=0.063. It was the failure of these statistics that led us to focus on the apparent successes with moles 6 and 9, and to the wording of the title of this paper. Nonetheless, owing to the small target size, the analysis in the previous paragraph tells us that the direct hits by moles 6 and 9 provide compelling evidence in favor of the hypothesis embedded in the title. This conclusion is bolstered by the second direct hit by mole 9, which was not included in our analysis.

The surface-walking moles (3 and 8) clearly were not attracted to the active seismic sources. This might be attributable to the spectral difference between the airborne sound emitted from the sand surface over our active seismic sources and the airborne sound emitted by *Stipagrostis* (the simulated vibrations from which we were using as stimuli). While the seismic stimulus components matched the *Stipagrostis* vibrational spectrum, the airborne stimulus components may have seemed alien to the mole subjects. The two surface-walking moles periodically head-dipped and thus were exposed to both the seismic and airborne components. While sand-swimming, on the other hand, the remaining eight moles were exposed only to the seismic component.

B. Implications regarding nonaural cues

The trail of mole 4 (left-hand panel of Fig. 4) branched twice and ended abruptly at two points (the sources that were active prior to battery failure are depicted in the figure). The trail began near the central clump of ostrich grass and eventually exited the arena close to the source at the north-northeast location. After exiting, it continued down the side of the ridge and off into a field of ostrich-grass clumps. Before exiting at that point, the mole took a long westward detour—along a route that passed directly over the source in the north-northwest position, then re-entered the arena and branched (northwest of the arena center) to form two blind ends. The mole clearly passed this branch point at least twice. Thus it was able to retrace at least part of its original path. It also must have passed the first branch point (north-northeast of the arena center) at least twice. It seems likely that the mole retraced the entire detour, bringing it back pre-

cisely to the original route. At least two cues could be available for this: (1) the topography of the sand surface along the trail, and (2) a scent-trail.

Near the beginning of its trail, mole 4 seems to have followed closely the route of mole 2, made approximately 24 h earlier. After it was measured and recorded, the trail of mole 2 had been raked thoroughly, and the wind subsequently had smoothed the surface of the sand. Nevertheless, it seems possible that mole 4 was attracted to a scent trail (or some other cues) left behind by mole 2. One sees a similar coincidence of the initial parts of the routes of moles 6 (male) and 10 (female). The laying down of those two trails was, however, separated by 6 days and at least two thorough rakings (of the trails of mole 6 and the trail of mole 7). Furthermore, other moles released in the arena during these six days did not follow the route of mole 6. The use of a scent-trail therefore seems unlikely.

For navigation in the absence of microseismic signals (e.g., under windless conditions, no nearby prey) golden moles presumably use nonaural cues, possibly based on tactile, olfactory, or even magnetic senses (Kimchi and Terkel, 2001). The apparent aversion of the sand-swimming moles to easterly routes may have resulted, in part, from a tendency of moles to follow routes close to those of their predecessors, utilizing such nonaural cues. On the other hand, moles 6 and 9 diverged strongly from the routes of their predecessors, taking routes directly toward active sources in the northwest quadrant. It should be noted that the arena was within a hundred meters of the eastern edge of the dune fields. An easterly path therefore, if continued, would take a mole out of its habitat.

C. Directional sensitivity

The results with moles 6 and 9 suggest that these moles were not only attracted toward the active sources, but were able to determine the directions of those sources. While in the sand-swimming mode, these moles were exposed to substrate-borne vibration but not (directly) to airborne sound. Mason and Narins (2002) briefly discussed three hypotheses for directional sensitivity to vibratory stimuli: (1) Use of interaural time or phase difference in the propagating seismic wave; (2) Directional sensitivity potentially imparted by a pushed-up ridge of sand acting as a seismic lens for surface-walking moles; (3) The localization potential of having nonparallel axes of rotation of the middle-ear ossicles. Use of interaural time or phase difference would require that a substantial component of the vibration imparted to the right-hand side of the mole's skull be independent of that imparted to the left side (i.e., that the skull not vibrate entirely as a unit). Such independence of motion awaits verification. The second hypothesis presumably would not apply to our sand-swimming moles. Here we expand on hypothesis 3.

Vertical particle motion, by itself, carries no information about the direction of the source. Such information would be carried, however, by a horizontal component of particle velocity that is aligned with the direction of seismic-wave propagation. In sand, seismic waves are propagated as surface (Rayleigh) waves or compressional (*P*) waves (Brownell, 1977). Appropriate horizontal particle-velocity components are present in both. At low frequencies, however, horizontal propagation of *P*-waves in sand may not be possible (Liu and Nagel, 1992). The vertical geophone in the arena monitored the vertical particle velocity component of the Rayleigh wave. The conduction velocity of Rayleigh waves in loose sand is approximately 40–50 m/s (Brownell, 1977), yielding a wavelength of approximately 80–100 cm at 500 Hz and greater than that for lower frequencies. The particles in a Rayleigh wave follow elliptical paths (for a graphical demonstration, see Russell, 2001). Within two-tenths of a wavelength of the surface, the horizontal motion at the top of the ellipse is directed toward the source; at the bottom of the ellipse it is directed away from the source. Below that depth, the elliptical motion is reversed. The sand-swimming moles were within 16 cm of the surface, making them well within two-tenths of a wavelength for the spectral components of our stimuli. By trial-and-error, seeking the direction of greatest stimulation, the mole could utilize the horizontal particle motion monaurally. The radiographs featured in Forster Cooper (1928) and Gasc, Jouffroy and Renous (1986) suggest that the axes of rotation of the mole's right and left ossicular chains are not parallel to each other in the horizontal plane. If that is so, then obliquely incident horizontal particle motion would excite the two ossicular chains to different extents, and the difference would bear information regarding source direction. This could allow binaural localization of the seismic source, an intriguing possibility that awaits experimental verification.

V. CONCLUSIONS

Statistical analysis of the data suggests, with only 93% confidence, that the paths of all eight sand-swimming were influenced by the vibratory fields generated by those sources. With confidences of 98% or 99% (depending on our choice of null hypothesis), on the other hand, the results imply that the paths of two of the eight sand-swimming moles were purposefully directed toward active sources, and that those two moles therefore must have used the sources as navigational beacons.

ACKNOWLEDGMENTS

The experiments and experimental procedures were approved by the Namibian Ministry of Environment and Tourism (Research/Collecting Permit No. 581/2002) and by the University of California, Berkeley, Animal Care and Use Committee (Protocol No. R087-1201). We are most grateful to Mike Griffin for facilitating this study, and to Marc and Elinor Durr for their gracious hospitality and assistance during our stay at De Duine Farmstead in the Namib-Rand Nature Reserve. We also are grateful to Hans Garüeb and Frans Haupundi for capturing the study animals. This research was supported in part by grants from the Committee on Research, University of California at Berkeley (E.R.L.), the UCLA Academic Senate and NIH Grant No. DC00222 (P.M.N.), and St. Catharine's College, University of Cambridge (M.J.M.).

APPENDIX: CALCULATION OF PROBABILITY

In this study, none of the eleven moles exited over the easterly 3/8 of the arena perimeter. In order to establish the likelihood that this was a chance occurrence, rather than the result of a westerly bias, we undertook the following analysis. Take the eight sources to divide the arena perimeter into a set of eight nonoverlapping segments of equal length, so that the first exit point of each mole can be assigned unambiguously to one of the eight. Assume that the *a priori* probability of its lying in a given segment is 1/8. Before the first trial, there are eight (overlapping) subsets that comprise three contiguous segments. After the first trial, three of those subsets are removed as candidates for the putative excluded zone. Unless they all produce first exit points in the same segment as the first, subsequent trials will reduce further the number of candidates. The goal here is to compute the probability that the number of candidates will be reduced to zero at or before the end of the eleventh trial. This can be accomplished by means of a seven-state Markov chain with the following stochastic projection matrix:

$$\mathbf{A} = [\mathbf{p_5}; \mathbf{p_4}; \mathbf{p_3}; \mathbf{p_{2b}}; \mathbf{p_{2a}}; \mathbf{p_1}; \mathbf{p_0}]$$

$$\mathbf{p_5} = [1/8,\ 1/4,\ 1/4,\ 1/4,\ 1/8,\ 0,\ 0]$$

$$\mathbf{p_4} = [0,\ 1/4,\ 1/4,\ 1/4,\ 0,\ 1/4,\ 0]$$

$$\mathbf{p_3} = [0,\ 0,\ 3/8,\ 1/4,\ 0,\ 1/4,\ 1/8]$$

$$\mathbf{p_{2b}} = [0,\ 0,\ 0,\ 1/2,\ 0,\ 1/4,\ 1/4]$$

$$\mathbf{p_{2a}} = [0,\ 0,\ 0,\ 0,\ 1/4,\ 3/4,\ 0]$$

$$\mathbf{p_1} = [0,\ 0,\ 0,\ 0,\ 0,\ 5/8,\ 3/8]$$

$$\mathbf{p_0} = [0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1].$$

This matrix operates on the vector $\mathbf{P}(n)$,

$$\mathbf{P}(n) = [P_5(n), P_4(n), P_3(n), P_{2b}(n), P_{2a}(n), P_1(n), P_0(n)],$$

where $P_m(n)$ is the probability that m three-segment candidates remain after n trials (there are two distinct states with two candidates remaining). Defining

$$\mathbf{P}(1) = [1,0,0,0,0,0,0],$$

one can carry the operation $\mathbf{A}\cdot\mathbf{P}$ through ten steps to compute $P_0(11)$. The probability that at least one candidate is left after 11 trials is its complement, $1-P_0(11)$.

$$P_0(11) = 0.958,$$

$$1 - P_0(11) = 0.042.$$

Autrum, H., and Schneider, W. (**1948**). "Vergleichende untersuchungen über den erschütterungssinn der insekten," (Comparative studies on the substrate-vibration sense of insects) Z. Vergl. Physiol. **31**, 77–88.

Brownell, P. H. (**1977**). "Compressional and surface waves in sand: Used by desert scorpions to locate prey," Science **197**, 479–482.

Brownell, P., and Farley, R. D. (**1979a**). "Detection of vibrations in sand by tarsal sense organs of the nocturnal scorpion, *Paruroctonus mesaensis*," J. Comp. Physiol. [A] **131**, 23–30.

Brownell, P., and Farley, R. D. (**1979b**). "Detection of vibrations in sand by tarsal sense organs of the nocturnal scorpion, *Paruroctonus mesaensis*: Mechanism of target location," J. Comp. Physiol. [A] **131**, 31–38.

Brownell, P., and Farley, R. D. (**1979c**). "Detection of vibrations in sand by tarsal sense organs of the nocturnal scorpion, *Paruroctonus mesaensis*: Orientation to substrate vibrations," J. Comp. Physiol. [A] **131**, 185–193.

Cocroft, R. B. (**2001**). "Vibrational communication and the ecology of group living, herbivorous insects," Am. Zool. **41**, 1215–1221.

Fay, R. R., and Popper, A. N. (**1985**). "The octavolateralis system," in *Functional Vertebrate Morphology*, edited by M. Hildebrand, D. M. Bramble, K. F. Liem, and D. B. Wake (Belknap Press, London), pp. 291–316.

Fielden, L. J. (**1991**). "Home range and movements of the Namib Desert golden mole, *Eremitalpa granti namibensis* (Chrysochloridae)," J. Zool. (London) **223**, 675–686.

Fielden, L. J., Hickman, G. C., and Perrin, M. R. (**1992**). "Locomotory activity in the Namib Desert golden mole, *Eremitalpa granti namibensis* (Chrysochloridae)," J. Zool. (London) **226**, 329–344.

Fielden, L. J., Perrin, M. R., and Hickman, G. C. (**1990**). "Feeding ecology and foraging behavior of the Namib Desert golden mole, *Eremitalpa granti namibensis* (Chrysochloridae)," J. Zool. (London) **220**, 367–389.

Forster Cooper, C. (**1928**). "On the ear region of certain of the Chrysochloridae," Philos. Trans. R. Soc. London, Ser. B **216**, 265–283.

Gasc, J. P., Jouffroy, F. K., and Renous, S. (**1986**). "Morphofunctional study of the digging system of the Namib Desert Golden mole (*Eremitalpa granti namibensis*): Cinefluorographical and anatomical analysis," J. Zool. (London) **208**, 9–35.

Gubbay, V. (**1956**). "A comparison of the development of the rudimentary eye of *Eremitalpa granti* (Broom) with that of the normal eye of *Elephantulus myrurus jameson* (Chubb)," South Afr. J. Science **52**, 182–186; **52**, 193–198.

Hartline, P. H. (**1969**). "Auditory and vibratory responses in the midbrains of snakes," Science **163**, 1221–1223.

Hartline, P. H. (**1971**). "Physiological basis for detection of sound and vibration in snakes," J. Exp. Biol. **54**, 349–371.

Hetherington, T. E. (**1989**). "Use of vibratory cues for detection of insect prey by the sandswimming lizard *Scincus scincus*," Anim. Behav. **37**, 290–297.

Hetherington, T. E. (**1992**). "Behavioral use of seismic cues by the sandswimming lizard *Scincus scincus*," Ethol. Ecol. Evol. **4**, 5–14.

Hickman, G. C. (**1990**). "The Chrysochloridae: Studies toward a broader perspective of adaptation in subterranean mammals," in *Evolution of Subterranean Mammals at the Organismal and Molecular Levels*, edited by E. Nevo and O. A. Reig (Wiley-Liss, New York), pp. 23–48.

Holm, E. (**1969**). "Contribution to the knowledge of the biology of the Namib Desert golden mole *Eremitalpa granti namibensis* (Bauer and Niethammer 1959)," Scient. Pap. Namib Desert Res. Stn. **41**, 37–42.

Kimchi, T., and Terkel, J. (**2001**). "Magnetic compass orientation in the blind mole rat, *Spalax elvenbergi.*," J. Exp. Biol. **204**, 751–758.

Kimchi, T., and Terkel, J. (**2002**). "Seeing and not seeing," Curr. Opin. Neurobiol. **12**, 728–734.

Koyama, H., Lewis, E. R., Leverenz, E. L., and Baird, R. L. (**1982**). "Acute seismic sensitivity in the bullfrog ear," Brain Res. **250**, 168–172.

Kuyper, M. A. (**1984**). "Golden moles," in *The Encyclopaedia of Mammals*, edited by D. MacDonald (George Allen and Unwin, London), pp. 764–765.

Lewis, E. R. (**1984**). "Inertial motion sensors," in *Comparative Physiology of Sensory Systems*, edited by L. Bolis and R. D. Keynes (Cambridge University Press, Cambridge), pp. 587–610.

Lewis, E. R., and Narins, P. M. (**1985**). "Do frogs communicate with seismic signals?" Science **227**, 187–189.

Lewis, E. R., Narins, P. M., Cortopassi, K. A., Yamada, W. M., Poinar, E. H., Moore, S. W., and Yu, X. (**2001**). "Do male white-lipped frogs use seismic signals for intraspecific communication?" Am. Zool. **41**, 1185–1199.

Liu, C. H., and Nagel, S. R. (**1992**). "Sound in sand," Phys. Rev. Lett. **68**, 2301–2304.

Lombard, R. E., and Hetherington, T. E. (**1993**). "Structural basis of hearing and sound transmission," in *The Skull*, edited by J. H. Hanken and B. K. Hall (University of Chicago Press, London), pp. 241–302.

Markl, H. (**1983**). "Vibrational Communication," in *Neuroethology and Behavioral Physiology*, edited by F. Huber and H. Markl (Springer-Verlag, Berlin), pp. 332–353.

Mason, M. J. (**2001**). "Middle ear structures in fossorial mammals: A com-

parison with non-fossorial species," J. Zool. (London) **255**, 467–486.
Mason, M. J. (**2003a**). "Bone conduction and seismic sensitivity in golden moles (Chrysochloridae)," J. Zool. (London) **260**, 405–413.
Mason, M. J. (**2003b**). "Morphology of the middle ear of golden moles (Chrysochloridae)," J. Zool. (London) **260**, 391–403.
Mason, M. J., and Narins, P. M. (**2001**). "Seismic signal use by fossorial mammals," Am. Zool. **41**, 1171–1184.
Mason, M. J., and Narins, P. M. (**2002**). "Seismic sensitivity in the desert golden mole (*Eremitalpa granti*): A review," J. Comp. Physiol. **116**, 158–163.
Meester, J. (**1964**). "Revision of the Chrysochloridae. 1. The desert golden mole *Eremitalpa* Roberts," Scient. Pap. Namib Desert Res. Stn. **26**, 1–18.
Narins, P. M., Lewis, E. R., Jarvis, J. U. M., and O'Riain, J. (**1997**). "The use of seismic signals by fossorial southern African mammals: A neuroethological gold mine," Brain Res. Bull. **44**, 641–646.
O'Connell-Rodwell, C. E., Hart, L. A., and Arnason, B. T. (**2001**). "Exploring the potential use of seismic waves as a communication channel by elephants and other large mammals," Am. Zool. **41**, 1157–1170.
Russell, D. (2001). "Acoustic animations.," http://www.gmi.edu/~drussell/Demos/waves/wavemotion.html
Seymour, R. S., Withers, P. C., and Weathers, W. W. (**1998**). "Energetics of burrowing, running and free-living in the Namib Desert golden mole (*Eremitalpa namibensis*)," J. Zool. (London) **244**, 107–117.
von Mayer, A., O'Brien, G., and Sarmiento, E. E. (**1995**). "Functional and systematic implications of the ear in golden moles (Chrysochloridae)," J. Zool. (London) **236**, 417–430.
Young, B. A., and Morain, M. (**2002**). "The use of ground-borne vibrations for prey localization in the Saharan sand vipers (*Cerastes*)," J. Exp. Biol. **205**(5), 661–665.

Hearing sensitivity and critical ratios of hooded crows (Corvus corone cornix)[a)]

Kenneth K. Jensen[b)] and Signe Klokker
Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

(Received 28 January 2005; revised 3 October 2005; accepted 28 November 2005)

The hearing threshold and critical ratios were estimated psycho-acoustically for captive wild-caught hooded crows by a yes/no procedure and the method of constant stimuli. Human subjects were tested in the same setup for direct comparison and to check for experimental artifacts. The hooded crows were found to have excellent low-frequency hearing capabilities compared to other passerine birds. Their hearing sensitivity is very close to that of humans at and below 5.6 kHz. The distribution of the critical ratios differed from that of the average bird and humans in being rather constant with frequency and not increasing monotonically. It furthermore showed a middle region of 5–6 dB lower critical ratio values between 500 Hz and 2 kHz. It is suggested that this improved range for hearing in noise is an adaptation to long distance communication. Human critical ratios gave the expected values and were between 3 and 6 dB lower than those of the crows. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2159431]

PACS number(s): 43.80−n, 43.66.Gf [JAS] Pages: 1269–1276

I. INTRODUCTION

Most Corvus species have a vocal repertoire called "structured cawing" typically consisting of three or more rhythmically repeated calls that are "harsh" quality, broadband, monotone signals (Goodwin, 1986; Laiolo and Rolando, 2003). Although it has never been demonstrated directly, there is little doubt in the literature that structured cawing plays an important role in territorial signaling (Coombs, 1978; Goodwin, 1986; Madge and Burn, 1994; Parr, 1997). Crows generally also have very large territories compared to other passerines (Goodwin, 1986). For the hooded and the carrion crow, territories of up to 0.5 km^2 in area, or 5–700 m in diameter, have been reported (Coombs, 1978). This is quite some distance compared to the typical 50–100-m cross section of the smaller song birds. Calling duels or "counter cawing" across territorial borders have been reported for the carrion, the hooded, and the American crow (Coombs, 1978; Goodwin, 1986) and certain call types in the hooded and carrion crow used in cawing duels have been reported audible at 700 m (Coombs, 1978).

Generally, for a *sender*, territorial signals are ideally aimed at individuals *outside* the territorial boundary. Signaling is adaptive because it saves the owner time and energy as compared to escalated aggressive behaviors (Bradbury and Vehrencamp, 1998). For *receivers* like territorial owners it ought to be adaptive to receive information about whether it is the well-known neighbor(s)—with whom the territory borders already have been settled—that are present or whether neighboring territories are open for incoming strangers. Furthermore, it must be adaptive to detect groups of nonbreeding predatory crows outside the territory, as they present a threat to eggs and nestlings (Coombs, 1978). For receivers like "floaters" (nonbreeding birds) it must be adaptive to obtain information on which areas are occupied as territories and to have a choice of involving themselves in conflict or not.

Altogether it thus seems reasonable to assume that the acoustic territorial communication of hooded crows ideally is adapted for functioning over distances proportional to their territorial sizes and that Corvus species therefore potentially represent an interesting case of long range acoustic communication.

Crows also appear worthwhile bird models for comparative studies to the human/mammalian hearing and perception. Few other studied bird species have vocalizations that are so closely comparable to the human voice and many other mammalian species as the crows. Crow calls consist of long duration, broadband, harmonic signals that apparently show formant structure and "vibratolike" modulations like in the human voice (Laiolo and Rolando, 2003). Rewarding contrasts and resemblances in fundamental auditory processing between crows and humans in their use and adaptations to similar communication signals might thus show up. In addition, comparing long distance communication in the crow to the general short distance communication in humans might help contrasting which elements of auditory processing and signal design are most important when the purpose is detecting faint, distant signals embedded in environmental noise compared to, for instance, speech intelligibility of louder signals, perhaps embedded in speech from other speakers at close range conditions.

Either way, it is important to start with the basics. In the present study we focus on two basic hearing abilities: hearing sensitivity threshold in the quiet and critical ratios using psychoacoustic methods on captive wild-caught hooded crows. We found that hooded crows had an excellent hearing

[a)]Portions of this work were presented at the 7th International Congress of Neuroethology, Nyborg, Denmark, August 2004.

[b)]Author to whom correspondence should be addressed. Electronic mail: kkj@biology.sdu.dk

sensitivity compared to other songbirds, and an interesting pattern of critical ratios suggesting an adaptation for detecting distant conspecific calls.

II. MATERIAL AND METHODS

A. Subjects

1. Hooded crows

Data were obtained from two wild-caught hooded crows, SP and CO, which were taken from the same nest in June 2001 at one month of age on the island of Falster in Denmark. Another two crows, BL and MB, added to the data where noted (both worked well, but BL died from disease and MB changed behavior and became too shy before finishing the experiments). BL was taken from a nest and MB caught as an adult in the university upland in June–July 2002. The sexes of the test birds were uncertain. The crows were kept outside in a flight cage $(3\times6\times2\frac{1}{2}\ \mathrm{m}^3)$. To motivate the crows for the psychoacoustic task, food access was restricted and the crows kept at 85%–90% of their free feeding weight.

2. Human subjects

Six human subjects were tested. They were all young listeners aged 24 to 31 years (including both authors). Four were women (SB, KJL, LST, and SK) and two were men (CB and KKJ). All had normal hearing sensitivity. Six subjects contributed to the hearing threshold measurements (SB, KJL, LST, SK, CB, and KKJ) and two contributed to the critical ratio measurements (SK and KKJ).

B. Technical setup

1. Test cages

Two identical test cages were built (so that two subjects could be tested at the same time and thereby save time). They had an outer layer of 13-mm gypsum boards supported by wooden frames. Three sides were covered with 100-mm and the other three with 200-mm mineral wool. The resulting inner dimensions of the test boxes were $105\times135\times95\ \mathrm{cm}^3$.

Inside and at one of the ends of each test cage was placed a test box consisting of a custom-made feeding delivery machine with a mounted test panel. The feeder delivered a single food pellet when activated (Vitakraft Beo Special). The test box's dimensions were $25\times20\times45\ \mathrm{cm}^3$ including a 2.5-cm layer of mineral wool at the top, the sides, and part of the front. The test panels themselves had the dimensions $15.5\times9.5\ \mathrm{cm}^2$ at the front and had three micro-switch activating pecking keys attached. Each key was 2 cm in diameter and they were separated by 4 cm (center to center). In the center of each key one LED was mounted: green in the left and right keys and red in the middle key. The distance from the cage floor to the center of the keys was 22–24 cm (variable due to the mineral wool floor). A speaker (see type below) was mounted in the roof of the test cage approximately 50 cm directly above the position of the experimental bird's head during test.

The test cages attenuated incoming sound by 35 dB at 250 Hz, increasing sharply to an attenuation of 50 dB at 500 Hz, and then increasing gradually to 60 dB at 8 kHz as measured by a frequency analyzer (HP 35665 A), a microphone (Brüel & Kjær, type 4133), and a loudspeaker (Nagra, Kudelski SA, machine 1491711). The overall ambient noise level inside the test cage was measured to 20 $\mathrm{dBA}_{\mathrm{eq}}$ or 62 $\mathrm{dBZ}_{\mathrm{eq}}$ (linear frequency weighting) (Brüel & Kjær sound level meter, type 2250). The overall ambient noise inside the lab in quiet was measured to 50 $\mathrm{dBA}_{\mathrm{eq}}$ or 78 $\mathrm{dBZ}_{\mathrm{eq}}$. But it has to be taken into account that these overall measures, including the A-weighted scale, are dominated by low frequencies (at and below approximately 2–300 Hz) and can only serve as a rough guide (the ambient noise level inside the test cages was too low to measure by 1/1 or 1/3 octave analysis).

2. Electronic setup

A Real Time Processor (Tucker-Davis, RP2.1) was used for tone stimulus generation by loudspeakers in each of the two test cages. Its two outputs were connected to two programmable attenuators (Tucker-Davis, PA5). The attenuator outputs were mixed (Monacor MMX-24) and connected to a stereo power amplifier (Tucker-Davis, SA1), which fed into a speaker (Vifa MG10MD09-08) in each test cage (see position above). The mixer was only necessary for the critical ratio experiments, but was used in both experiments so that the only difference between the two experiments was the addition of noise. A web cam was mounted on the end wall, opposite to the test panel, to enable the experimenter to observe and record the birds' behavior during experiments.

The electronic setups for both the hearing threshold and the critical ratio experiments were identical, except for the addition of noise by a Hewlett Packard Noise Generator (3722A) in the latter ones. Its two outputs were connected to each of two graphic equalizers (*dbx* type 131) and then mixed with the tone signal. The whole system, stimulus generation, stimulus level control, key-peck registration, cage lights, feeding automate, etc., was computer controlled by custom-made software.

C. Test procedure on crows

1. Sound stimuli and calibration

The test sounds were sinusoids of 400-ms duration including a 5-ms rise and fall time ($\cos^2$-gate) and repeated in pulses separated by 30 ms (cf. Dooling and Okanoya, 1995). This signal duration is a factor of 1.5 to 2 above the maximal avian temporal integration time (Dooling *et al.*, 2000). The sound level was calibrated at least once a day with the microphone in a position normally occupied by the bird's head during testing ($\frac{1}{2}$ in. Brüel & Kjær type 4176 microphone connected to a Brüel & Kjær type 2235 sound level meter with settings "linear weighting" and "random incidence").

For all the different test frequencies the variation in the sound level was measured at six different positions by displacing the microphone 2 cm forwards, backwards, upwards, downwards, and to either side relative to the calibration position. This was the spatial matrix, where the crows' head/

ears would typically be located during testing. Of these measurements, 87% deviated by up to 2 dB relative to the calibration position and the remaining 13% by up to 3.5 dB (mostly at frequencies between 2 and 4 kHz). It should be added that each crow remained in a rather constant, fixed position in front of the test panel from where it could reach all three keys and the crows did not move their heads to "sample" the sound field during the time before responding.

During critical ratio experiments, we added white noise in the range from 0 to 20 kHz and used the equalizers to adjust the noise floor power spectrum to be flat within 5 dB at the calibration position. The power spectrum density of the noise was adjusted to 1.5 ± 0.5 dB/Hz (*re* 20 μPa^2) in the range from 0 to 12.5 kHz. Critical ratios were calculated by subtracting the power density level from the measured masked threshold.

2. Psychoacoustic procedure and threshold determination

We used a *yes/no procedure* of the type known as "*go left/go right*" in the avian literature (Hulse, 1995) together with the *method of constant stimuli* with five constant sound levels separated by a 3-dB step size (see, e.g., Hienz *et al.*, 1977). The middle tone level was adjusted to lie around the assumed threshold before session start. *Tone* or *no tone* was presented by an overall 50% probability.

In our experimental setup, each crow learned to initiate a trial of either *tone* or *no tone* by pecking the middle key and then proceeded to peck the right key if it heard a sound (respond "yes") or the left key if it did not (respond "no"). Pecks at the right key were thus registered as *hit* if a tone was in fact presented and as *false alarm* if no tone was presented. Pecks at the left key were registered as *miss* if a tone was in fact presented and *correct rejection* if no tone was presented. Correct responses (hit/correct rejection) were food rewarded and wrong responses (FA/miss) were punished by a 10–20-s blackout of cage illumination depending on the individual. For any given tone level the hit proportion was calculated as the number of registered hits at this tone level divided by the total number of presentations in a session. The proportion of false alarms was calculated as the number of registered false alarms divided by the total number of no tone presentations in a session.

In order to adapt the crows' behavior to the overall procedure in practice and to facilitate data collection the following four subprocedures were implemented:

(1) In order to avoid too many similar trials in a row and a potential temporary build up of key preference no more than *three successive* tone or no tone trials were allowed in a row (Klump, 1995).
(2) To assure that each of the five tone levels had been presented an approximately equal number of times during a session, the tone levels were arranged into *sequential blocks* of 15 tone level stimuli, within which each of the five tone levels was represented three times. Tone levels were picked randomly from such a block during any tone trial, and a given block had to be "emptied" before the next could start.
(3) To keep the subjects *conservative* and under *stimulus control* and at the same time be able to run *more trials* in a session, the crows were required to make two consecutive correct responses to get a food reward (Klump, 1995). Thus, if the crows pecked the keys at random, they would only receive a reward with a 25% chance (and on average wait for three times the duration of the blackout time before it received it). The food delivery machine was always activated following correct responses so that the sound would function as a secondary positive reinforcer. Punishment was delivered following *every* wrong response.
(4) Lastly we implemented *corrections trials* (Hienz *et al.*, 1977) in the procedure such that in case of a miss the next trial presented would be the loudest tone, while in case of a FA a no tone trial was presented. This assured that no key would have to be pecked markedly more often than the other one, avoiding potential temporary buildup of key preference. At the same time correction trials served as reminders of the procedure and in case of a *tone correction trial* also as a stimulus reminder. Results from correction trials were not included in the analysis.

Using these procedures, the crows were allowed to operate the panel until they were satiated. If such a session contained 200 trials or more, it was used for threshold determination.

After the end of a session we calculated d' values from the hit percentages and the false alarm rate, and plotted d' as a function of tone level. Calculations of d' for each test sound pressure level were done by subtracting the standard ($\mu = 0, \sigma = 1$) inverse normal distribution score of the FA proportion, Z(FA proportion), from the standard inverse normal distribution score of the hit proportion, Z(hit proportion) (see, e.g., Green and Swets, 1974). Our threshold criterion was chosen at $d' = 1$ (one standard deviation) and we read the threshold off the plot at this value by linear interpolation.

For the response procedure, we used a *deferred decision design* (Birdsall and Roberts, 1965; Swets and Birdsall, 1967), in which the observer is allowed to defer his decision until an optional number of observations have been made (see below). Thus the allowed number of observations or response time after initiating a trial was unlimited and the crows could use whatever time they wished to decide their responses. We did this because it seemed to function best with the crows' behavior in that they tended to get very frustrated and unstable in their response behavior when they were cut off by a limited response time, and the alternative of an unlimited response time following a single tone presentation might pose a problem with stimulus control. The deferred decision design meant that the experimentally obtained thresholds had to be corrected for the increase in detectability produced by the repetition of the signal. That is, we had to calculate what the experimentally obtained threshold would be, if only one observation had been available. Thus the experimental threshold was corrected to a threshold value corresponding to only *one* observation according to *deferred decision theory* and *the integration model* (Birdsall

and Roberts, 1965; Green and Swets, 1974; Swets and Birdsall, 1967).

Although this theory is well described in the literature, we will describe it here in some detail, since this is the first time the theory has been used in the avian literature. First of all, experimental tests in humans show a good agreement with the integration model (Swets *et al.*, 1959, 1978; Swets and Birdsall, 1967) and we assume the same holds true for hooded crows. According to the integration model, all information is integrated and the detectability will change as follows (Green and Swets, 1974):

$$d'_n = \sqrt{\sum_{i=1}^{n} (d'_i)^2}, \tag{1}$$

where d'_n is the combined detectability, d'_i the detectability of each observation, and n is the total number of observations. That is, the detectability d'_n, combined from multiple (n) observations, is equal to the root of the sum of the squares of the individual d' values, d'_i, associated with every single observation. It is assumed that all d'_i are equal and since we have chosen to read off the experimentally obtained threshold at $d'=1$ and since this d' is the combined detectability equal to d'_n we get (Green and Swets, 1974)

$$d'_i = \frac{d'_n}{\sqrt{n}} = \frac{1}{\sqrt{n}}. \tag{2}$$

Since detectability is approximately proportional to signal power (Swets *et al.*, 1959), we can then calculate what the corrected threshold ($\text{threshold}_{\text{DDT}}$) would be in a given session in relation to the actually obtained experimental threshold ($\text{threshold}_{\text{EXP}}$) by the following equation (Swets *et al.*, 1959):

$$\text{threshold}_{\text{DDT}} = \text{threshold}_{\text{EXP}} - 10\log\left(\frac{1}{\sqrt{\bar{n}_{\text{THR}}}}\right), \tag{3}$$

where $\bar{n}_{\text{THR}}$ is the mean number of observations made before responding at the experimental threshold, $\text{threshold}_{\text{EXP}}$. The mean numbers of observations were calculated from response latencies. Response latencies were software recorded as the time difference in ms between time of trial initiation (the subject presses the middle button) and cessation (the subject responds by pressing the right or left button). To calculate $\bar{n}_{\text{THR}}$, we first had to find the mean response time at $\text{threshold}_{\text{EXP}}$. That was done by averaging the response times for each of the five tone levels and plotting these mean values of response times as a function of tone level. The mean response time at $\text{threshold}_{\text{EXP}}$ was then read off the plot by linear interpolation. Then, by assuming that each single 400-ms tone pulse constituted *one* observation and since a tone burst was followed by 30-ms silence, the average number of observations, $\bar{n}_{\text{THR}}$, was estimated as

$$\bar{n}_{\text{THR}} = \frac{\text{response time at threshold}_{\text{EXP}}\,(\text{ms})}{430\ \text{ms}}. \tag{4}$$

3. Test frequencies and chronology

The hearing thresholds were tested in $\frac{1}{2}$-octave steps from 250 Hz up to 8 kHz. The four test birds were trained at 2.5 kHz before the experimental period. They were generally tested twice a day, five days a week from June to September 2003. When a hearing threshold had been determined at any one given test frequency, the critical ratio at the same frequency was measured in sessions directly following. Thus the two different experiments were run overlapping in time. The crows were first tested stepwise in the $\frac{1}{2}$ octave steps from 2.5 kHz down to 250 Hz and then stepwise in the $\frac{1}{2}$ octave steps from 2.8 kHz up to 8 kHz. The final threshold at any given frequency was determined as the average of three successive sessions that were within a 3-dB range and did not successively decrease.

D. Test procedure on humans

Human subjects were tested at four frequencies: 0.35, 1, 2.8, and 5.6 kHz. The four frequencies were assigned to the test persons as follows: SB and KJL=1 and 2.8 kHz; LST and CB=350 Hz and 5.6 kHz; and SK and KKJ did all four frequencies, such that $N=4$ at all four test frequencies. In the critical ratio experiment both subjects (SK and KKJ) did all four frequencies. Subjects were instructed to sit in a position, with their forehead position fixed by resting it on the feeder box wall just above the key panel. They were instructed to keep this position as symmetrical and constant as possible during the session. The head was tilted forwards 45° to 60° in the listening position, so that the sound waves would be perpendicular to the surface of the back of the head (the sound coming directly from above as in the crows). The sound level was calibrated 5 cm away from the meatus at both ears and the highest value of the two was used as the calibration reference (measurements never differed more than 2.5 dB SPL). Before critical ratio experiments, the noise floor power spectrum was adjusted to be flat within a 5-dB range from 0 to 20 kHz at the ear with the maximal calibration value. Here the power spectrum density of the noise was adjusted to 1.5 ± 0.5 dB/Hz (*re* 20 μPa^2) in the range 0–20 kHz. The human subjects were assigned test frequencies at random but the tests were otherwise run in the exact same manner as in birds except for the following changes: (1) hearing threshold experiments were fully completed before critical ratio experiments were performed, (2) the number of trials per session was reduced to 100, (3) we omitted the subprocedure allowing only three similar trials in a row, and (4) we kept blackout time to 2 s.

III. RESULTS

The crow data are predominantly from individuals SP and CO, which completed all test frequencies in both experiments. BL and MB ended up supplying data for one frequency each (2 and 4 kHz, respectively) in the hearing threshold experiments in addition to the training frequency at 2.5 kHz. MR also produced data at 2 kHz in the critical ratio experiments. SP, CO, and MB had a blackout time of 20 s, whereas BL had one of 10 s. False alarm rates under this condition, for all crows in both experiments, generally re-

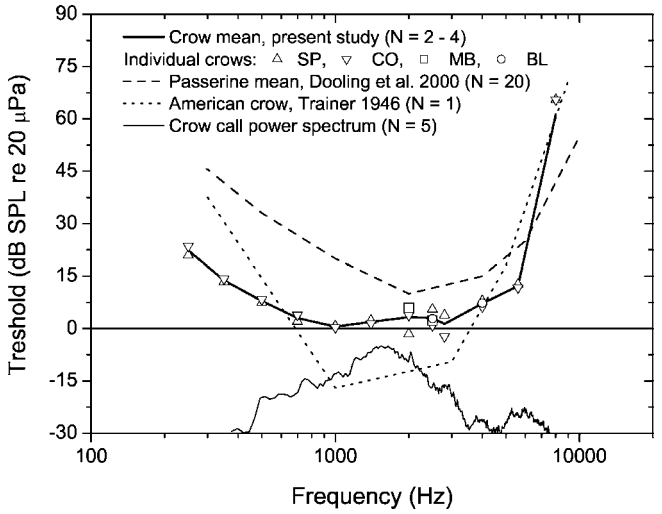


FIG. 1. Hearing thresholds for wild caught hooded crows. The mean passerine hearing threshold and the hearing threshold of one specimen of the American crow are shown for comparison. Also shown is an average power spectrum of hooded crow calls recorded in the field at short range [natural open field habitat, distance <20 m, Sony DAT recorder (TCD-D10 Pro) and a handheld Beyer Dynamics shotgun microphone (MC737 PV)].

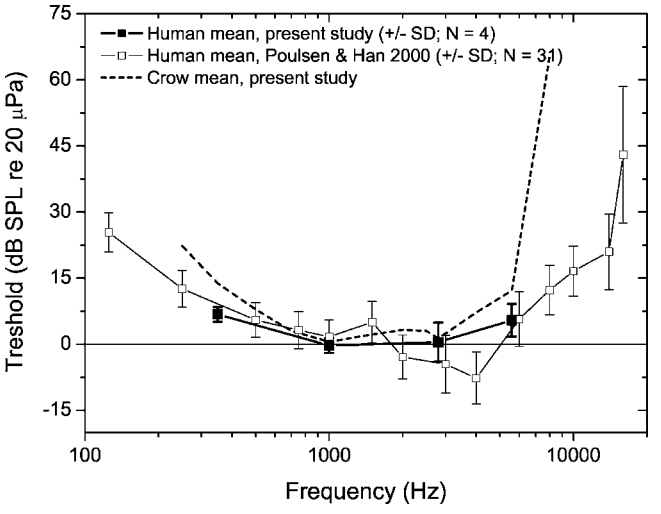


FIG. 2. Hearing thresholds of human subjects. An earlier free-field study on human subjects and the present study hearing threshold of the hooded crows are shown for comparison.

mained between 10% and 20%, but ranged from 0% to 30%. The human observers received, as specified above, a symbolic blackout time of 2 s. Even in this condition they were generally very conservative with false alarm rates near 5%, but extending up to 15%.

A. Hearing threshold

The hooded crow average response latency at threshold for all frequencies was 1788±625 ms (SD) corresponding to 4.2±1.5 presentations (SD). We excluded crow subject MB from this average because its latency responses [mean =4,036±308 ms (SD)] were more a measure of its extreme cautiousness than of its detection process. The average correction factor for multiple observations [cf. Eq. (3)] and thus the average difference between uncorrected and corrected thresholds was 3.1±0.8 dB (SD). The crows had their best hearing in the range between 700 Hz and 2.8 kHz, where the threshold was a few dB above 0 dB SPL (Fig. 1). From the best hearing range and towards lower frequencies, the hearing sensitivity decreased at approximately 12–14 dB per octave, whereas, towards higher frequencies, the hearing sensitivity decreased much steeper by approximately 30–40 dB per octave.

The human average response latency for all individuals was 1564±194 ms (SD) corresponding to 3.6±0.5 presentations (SD). The average correction factor for multiple observations was 2.8±0.3 dB (SD). The human subjects showed a best hearing sensitivity at 1 and 2.8 kHz around 0 dB SPL (Fig. 2). At 0.375 and 5.6 kHz the hearing sensitivity had decreased by approximately 7 and 5 dB SPL, respectively.

B. Critical ratios

The hooded crow average response latency at threshold for all frequencies was 1520±745 ms (SD) or 3.5±1.7 observations (SD) (again crow subject MB was excluded [response latency=3408±483 ms (SD)]). The average correction factor for multiple observations [Eq. (3)] was 2.7±1.2 dB (SD). The hooded crow critical ratios seemed to fall in three spectral ranges: a low-frequency range from about 250 to 500 Hz with critical ratios up at 24–28 dB, a mid-frequency range from 750 Hz to 1.4 kHz with the critical ratios down to approximately 21–22 dB, and a high-frequency range from 2.5 to 5.6 kHz with critical ratios back up at approximately 27–28 dB (Fig. 3).

The human average response latency for all individuals was 1098±140 ms (SD) or 2.6±0.3 observations (SD). The average correction factor for multiple observations was 2.0±0.2 dB (SD). In contrast to the crows, the average human critical ratios increased with frequency from approximately 17 dB at 0.375 kHz to approximately 21–22 dB at 2.8 and 5.6 kHz (Fig. 3).

IV. DISCUSSION

A. Hearing threshold

Our data show that the hooded crow has an excellent overall hearing sensitivity, especially in the low-frequency

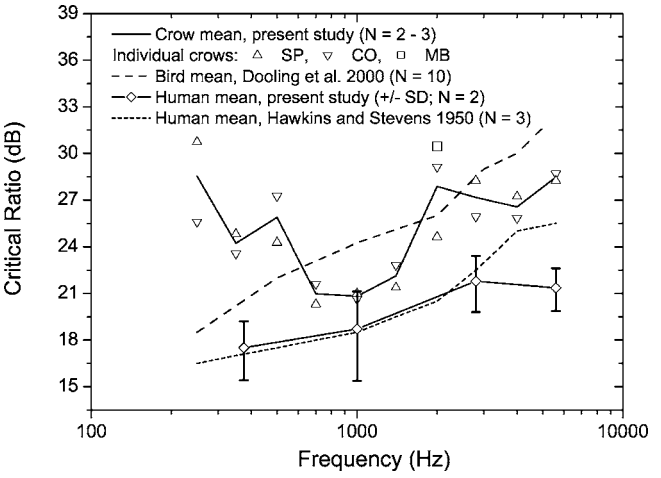


FIG. 3. Critical ratios of hooded crows and human subjects. The bird mean critical ratios and an earlier result on human subjects are shown for comparison.

range (<2 kHz) compared to that of the average passerine bird (Fig. 1). As most data were obtained from only two sibling crows, generalizations must be made with caution. However, two observations suggest that our findings are close to the general picture. First, the overall distribution of our data points fits within a general distribution of many other vertebrates, especially mammals, with comparable low-frequency hearing sensitivities shown in Fay (1988), like, for instance, humans and chimpanzees (*Pan troglodytes*), p. 327; squirrel monkeys (*Saimiri sciureus*), p. 335; domestic cat (*Felis catus*), p. 347; dog (*Canis canis*), p. 353, and many others. Second, the few data points from unrelated individuals did not deviate notably from the overall distribution (standard deviation still low).

The comparatively good sensitivity at low frequencies is not surprising, since larger size and weight have been shown to be positively correlated with low-frequency sensitivity in birds (Rosowski and Graybeal, 1991), and hooded crows are rather large compared to the average passerine bird. Their greater size also makes them able to produce low-frequency vocalizations more efficiently (Fletcher, 2004) and the power spectrum of their vocalizations (recorded at short range) closely matches their hearing sensitivity (Fig. 1). Lower frequencies generally transmit better in the environment than higher frequencies (Wiley and Richards, 1982). Consequently, hooded crows seem well suited for long distance acoustic communication.

The hearing sensitivity of two other corvine species have been tested previously: the American crow (Trainer, 1946) and the blue jay (*Cyanocitta cristata*) (Cohen *et al.*, 1978). The hearing threshold of the American crow has been found to be as low as −17 dB SPL at 1 kHz (Fig. 1). This is the lowest threshold recorded in any passerine and is as low as the best frequency of many owl species (see, e.g., Dyson *et al.*, 1998). Trainer (1946) carried out his early experiments with great care and thoughtfulness. Many of his results in other species are in general comparable to more recent replications and are commonly referenced. He determined an equally low threshold measurement of −16 dB SPL at 1 kHz in the great horned owl (*Bubo virginianus*). This is in close agreement with recent measurements on other owl species (Dyson *et al.*, 1998) and indicates that his measurements of such low threshold values seem in general valid. Even though there are many discrepancies in the experimental setup between our and Trainer's experiments, no single factor, nor a sum of multiple factors, can possibly explain a discrepancy of −17 dB (SPL) in threshold, not even the negative reinforcement of electric shock used by Trainer versus our positive food reinforcement scheme. Thus, either the hearing sensitivity of the American crow is in general significantly lower than that of hooded crows around 1 kHz, or the single individual used by Trainer had an atypical considerable increased hearing sensitivity. The other possibility is that the two hooded crows in the present study, caught as nestlings and thereby probably siblings, could oppositely have an inherent, atypical *decreased* hearing sensitivity at 1 kHz. But since our data include two to four individuals and follow in overall form both the average passerine, and many other animal hearing thresholds, this seems to be less likely.

The hearing threshold curve of the blue jay (Cohen *et al.*, 1978) shares more similarities with that of the average passerine bird than with the two Corvus species and does not exhibit the high sensitivity at low frequencies. Again, this is most probably related to body proportions (Rosowski and Graybeal, 1991): the blue jay is comparable to, for instance, the European starling (*Sturnus vulgaris*), both in body weight and size (Cramp and Perrins, 1994; Madge and Burn, 1994) and in its hearing sensitivity (Fay, 1988).

The human thresholds were nearly identical with earlier results like, for instance, those of Poulsen and Han (2000) (Fig. 2). This indicates that our psychoacoustic setup is functionally not different from others at the tested frequencies and, apparently, that no experimental artifacts have influenced the results on the hooded crows. The human results also show by direct comparison that the hearing sensitivity of hooded crows is comparable to that of humans at and below 5.6 kHz except around 4 kHz (Fig. 2) where human hearing is facilitated by pinna and meatus resonance (Moore, 2003).

B. Critical ratios

Due to the apparent unique pattern of the critical ratios these results are more difficult to interpret. The results from the two human subjects were again well within expected results. Thus by direct comparison this suggests that experimental artifacts are unlikely to explain the atypical pattern. This also means that the only obvious factor which could be different between the tests of human and crow subjects is the sound field and its variation around their ears. This is unlikely to play a role though, since (1) no sound field variation measured for the crows was large enough to account for a 5–6-dB steep change from 1.4 up to 2 kHz or from 500 down to 750 Hz, and (2) no such relative large increment was found in the hearing threshold data, which would be expected to be affected to the same degree as the critical ratios. Thus, altogether, experimental artifacts seem to be excluded.

This leaves out the following two most likely explanations that either the two hooded crows, probably being siblings, had inherent divergent hearing abilities, or this special critical ratio pattern is a general hooded crow trait. Since the hearing thresholds of the two hooded crows, both qualitatively and quantitatively, were within the expected range and form relative to their size compared to many other birds and mammals (see above), and in general were very convincingly distributed, without any great fluctuations between neighboring frequency measurements, we see no reason to believe that the two hooded crows should have any significant deviating hearing abilities.

In conclusion, we thus interpret our findings such that in the frequency range between approximately 500 Hz and 2 kHz the hooded crow is specially adapted for hearing in noise. At 1 kHz, for instance, the critical ratio was approximately 3.5 dB lower than that of the average bird (Dooling *et al.*, 2000) and indicates an improved ability to hear in noise around this frequency. It needs to be established, though, whether the observed pattern of critical ratios is conserved at noise levels other than that with a power spectrum density of

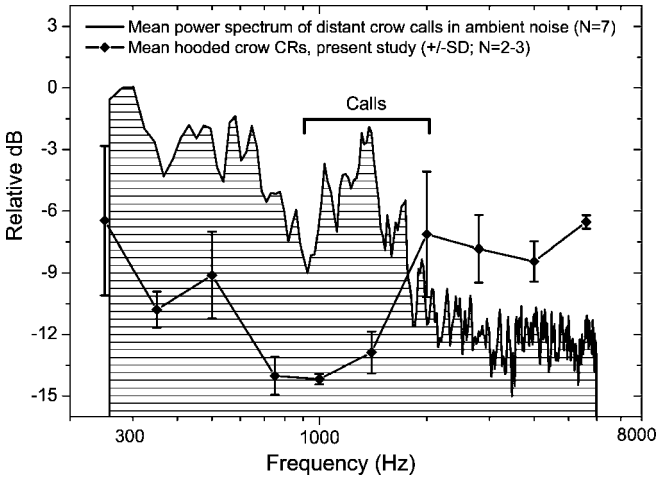


FIG. 4. Comparison of the hooded crow critical ratio pattern to the mean power spectrum of hooded crow calls recorded in the field at a long distance [natural open field habitat, distance above at least 300 m, Sony DAT recorder (TCD-D10 Pro) and a handheld Beyer Dynamics shotgun microphone (MC737 PV)].

1.5 dB/Hz (*re* 20 μPa^2) used in the present study. It is reasonable to assume that the pattern is level independent since critical ratios have been experimentally shown to be constant at power spectrum densities between 1 and 41 dB/Hz (*re* 20 μPa^2) in European starlings (*Sturnus vulgaris*).

1. Adaptation for detecting distant crow calls?

When comparing the range of low critical ratios to long distance transmitted calls (>300 m; N=7), all the spectral energy (not completely covered by natural noise) lies within this range (Fig. 4), suggesting that the low critical ratio range is an adaptation for hearing out distant crow calls. The spectral energy lies entirely at the high-frequency half of the low critical ratio range though and not in the middle as might be expected if the low critical ratio range was to be such an adaptation. However, this might be explained by *spread of masking* (Moore, 2003): from higher frequencies (>2 kHz) much spread of masking down towards the crow signal is not as likely, and the signal might then as well be placed up towards the high-frequency limit of the low critical ratio range. On the other hand, upward spread of masking from the lower frequencies is known to be much more pronounced, at least in mammals (Moore, 2003), and the lower frequency half of the low critical ratio range not covering the signal might be necessary for improved resolution of low-frequency noise and coping with upwards spread of masking. If this is the case, and since 500 Hz is the low-frequency limit of the low critical ratio range, we would expect approximately 500 Hz to be the frequency limit, above which upwards spread of masking is having a significant masking effect on frequencies between 1 and 2 kHz within natural noise levels. To our knowledge, no direct measurement of spread of masking has been performed on birds. However, the more symmetrical psychophysical tuning curves (Dooling *et al.*, 2000) and avian auditory nerve fiber tuning curves (Gleich and Manley, 2000) might indicate that upwards spread of masking may not be as prominent in birds as in mammals. Nevertheless, natural noise is mostly much more pronounced in the low-frequency end (see, e.g., Fig. 4) and masking from this region is more likely to occur.

2. Diverging critical ratio patterns in general

The general pattern in mammalian and avian species is that the critical ratios increase monotonically with approximately 3 dB per octave, but with some variation (see, e.g., Fay, 1988). Critical ratios distributed notably differently from this monotonic 3-dB increase have been observed in a number of other birds like the budgerigar (*Melopsittacus undulatus*) (see references in Dooling *et al.*, 2000), the orange-fronted conure (*Aratinga canicularis*) (Wright *et al.*, 2003), the great tit (*Parus major*) (Langemann *et al.*, 1998), and North American "blackbirds" [pooled data from red-winged blackbirds (*Agelaius phoeniceus*) and brown-headed cowbirds (*Molothrusater*)] (Hienz and Sachs, 1987). Critical ratios are to some degree seen as an indirect measure of frequency selectivity and of at least *relative* sizes between auditory filter bandwidths (see, e.g., Moore, 2003). The precise origin of the auditory filters is still somewhat uncertain although the mechanical properties basilar membrane (Moore, 2003), or the analog basilar papilla in birds (Manley, 1990), is clearly involved. A number of structures and mechanisms are potentially influencing the frequency selectivity of the basilar papilla, such as its mechanical properties, electrical tuning of hair cells, mechanical properties of the stereovillar bundle, and its interaction with the tectorial membrane and basilar papilla (Manley, 1990). Morphological studies have not revealed any obvious features in the peripheral hearing organ that could explain the existence of the critical ratio pattern in the budgerigar, which is rather constant with frequency and thus remains low at higher frequencies (Manley *et al.*, 1993). We suggest that the hooded crow would be a better experimental animal than the budgerigar to elucidate morphological and mechanical adaptations in the auditory periphery since the three different critical ratio regions, indicating regions of different frequency selectivity, might show up as three different contrasting regions in structural or biochemical characteristics.

C. Procedural issues

The deferred decision design where the crows and human subjects were allowed an unlimited number of tone presentations and response time generally meant a minor adjustment of 2–3 dB SPL upwards of threshold values. The test results on human hearing sensitivity and critical ratios were very close to those of other studies. This suggests that our experimental procedures including the deferred decision design functioned well and that the results do not seem to differ significantly from other procedures and designs. The deferred decision design allows subjects to calmly listen and decide on their response. In addition, the results from the human subjects were very close to the established ones despite a relatively low N and tone trial repetitions. We therefore are under the impression that this design minimizes unwanted procedural statistical noise and variation. This of course needs experimental verification.

Finally, we want to mention that many Corvus species repeat their monotone, typically 300- to 400-ms duration, call rhythmically a number of times (Goodwin, 1986; Madge and Burn, 1994), which for the hooded crow is typically three to five times (personal observations). According to the deferred decision theory, such repetitions will lower the detection threshold relative to that of a single call by a factor of (number of call repetitions)$^{0.5}$. In the case of four repetitions, for instance, the detection threshold would be reduced by 3 dB.

V. CONCLUSION

The hooded crow has an excellent low-frequency hearing compared to other passerines, which correlates well with its low-frequency vocalizations. This is probably attributed to their relative large size. Together with the fact that lower frequencies in general transmit best, hooded crows are well suited for long distance communication. In addition, hooded crows show an interesting atypical pattern of critical ratios, which seems to be an adaptation for detecting distant crow calls. We furthermore conclude by direct comparison that hooded crow hearing sensitivity at and below 5.6 kHz is very comparable to that of humans, while the critical ratios show a quite different pattern.

ACKNOWLEDGMENTS

The present study was performed with permissions and counseling from the Danish Ringing Centre, the Danish Forest and Nature Agency, and the Danish Animal Experiments Inspectorate. We thank gamekeeper Benny Larsen for advice on catching crows, Christian Brandt, Signe Brinkløv, Katrine Juul Larsen, Louise S. Tjørnum, and Claes U. J. Hansen for willingness to act as experimental subjects, and Frank Mortensen, Claus Poulsen, and Robert D. Hienz for technical and procedural help and advice. We are grateful to Jakob Tougaard, Kristian Beedholm, Ole Næsbye Larsen, and two anonymous reviewers for valuable comments on earlier versions of the manuscript. Experiments with human subjects were performed with permission from the Danish National Committee for Biomedical Research Ethics (region Vejle & Fyns amt). Work was supported by a grant from the Danish National Research Foundation to the Center for Sound Communication.

Birdsall, T. G., and Roberts, R. A. (**1965**). "Theory of signal detectability—deferred-decision theory," J. Acoust. Soc. Am. **37**, 1064–1074.

Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderland, MA).

Cohen, S. M., Stebbins, W. C., and Moody, D. B. (**1978**). "Audibility thresholds of the blue jay," Auk **95**, 563–568.

Coombs, C. J. F. (**1978**). *The Crows: A Study of the Corvids of Europe* (Batsford, London).

Cramp, S., and Perrins, C. M. (**1994**). *Handbook of the Birds of Europe, the Middle East, and North Africa: The Birds of the Western Palearctic—Volume VIII: Crows to Finches* (Oxford U.P., Oxford).

Dooling, R. J., and Okanoya, K. (**1995**). "Psychophysical methods for assessing perceptual categories," in *Methods in Comparative Psychoacoustics*, Vol. **6**, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser Verlag, Basel, Switzerland), pp. 307–318.

Dooling, R. J., Lohr, B., and Dent, M. L. (**2000**). "Hearing in birds and reptiles," in *Comparative Hearing: Birds and Reptiles*, Vol. **13**, edited by R. J. Dooling, R. R. Fay, and A. N. Popper (Springer, New York), pp. 309–359.

Dyson, M. L., Klump, G. M., and Gauger, B. (**1998**). "Absolute hearing thresholds and critical masking ratios in the European barn owl: a comparison with other owls," J. Comp. Physiol., A **182**, 695–702.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay, Winnetka, IL).

Fletcher, N. H. (**2004**). "A simple frequency-scaling rule for animal communication," J. Acoust. Soc. Am. **115**, 2334–2338.

Gleich, O., and Manley, G. A. (**2000**). "The Hearing Organ of Birds and Crocodilia," in *Comparative Hearing: Birds and Reptiles*, edited by R. J. Dooling, R. R. Fay, and A. N. Popper (Springer, New York), pp. 70–138.

Goodwin, D. (**1986**). *Crows of the World* (British Museum of Natural History, London).

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, Huntington, NY).

Hienz, R. D., and Sachs, M. B. (**1987**). "Effects of noise on pure-tone thresholds in blackbirds (*Agelaius-phoeniceus* and *Molothrus-ater*) and pigeons (*Columba livia*)," J. Comp. Psychol. **101**, 16–24.

Hienz, R. D., Sinnott, J. M., and Sachs, M. B. (**1977**). "Auditory-sensitivity of redwing blackbird (*Agelaius-phoeniceus*) and brown-headed cowbird (*Molothrus ater*)," J. Comp. Physiol. Psychol. **91**, 1365–1376.

Hulse, S. H. (**1995**). "The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals," in *Methods in Comparative Psychoacoustics*, Vol. **6**, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser Verlag, Basel, Switzerland), pp. 319–330.

Klump, G. M. (**1995**). "Sound localization studies in non-specialized birds," in *Methods in Comparative Psychoacoustics*, Vol. **6**, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhäuser Verlag, Basel, Switzerland), pp. 171–182.

Laiolo, P., and Rolando, A. (**2003**). "The evolution of vocalizations in the genus Corvus: effects of phylogeny, morphology and habitat," Evol. Ecol. **17**, 111–123.

Langemann, U., Gauger, B., and Klump, G. M. (**1998**). "Auditory sensitivity in the great tit: perception of signals in the presence and absence of noise," Anim. Behav. **56**, 763–769.

Madge, S., and Burn, H. (**1994**). *Crows and Jays: A Guide to the Crows, Jays and Magpies of the World* (A & C Black, London).

Manley, G. A. (**1990**). *Peripheral Hearing Mechanisms in Reptiles and Birds* (Springer-Verlag, Berlin).

Manley, G. A., Schwabedissen, G., and Gleich, O. (**1993**). "Morphology of the basilar papilla of the budgerigar, *Melopsittacus undulatus*," J. Morphol. **218**, 153–165.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, San Diego).

Parr, C. S. (1997). "Social Behavior and Long Distance Vocal Communication in Eastern American Crows," thesis/dissertation, The University of Michigan.

Poulsen, T., and Han, L. A. (**2000**). "The binaural free field hearing threshold for pure tones from 125 Hz to 16 kHz," Acustica **86**, 333–337.

Rosowski, J. J., and Graybeal, A. (**1991**). "What did Morganucodon hear?" Zool. J. Lin. Soc. **101**, 131–168.

Swets, J. A., and Birdsall, T. G. (**1967**). "Deferred decision in human signal detection—preliminary experiment," Percept. Psychophys. **2**, 15–28.

Swets, J. A., Green, D. M., Getty, D. J., and Swets, J. B. (**1978**). "Signal-detection and identification at successive stages of observation," Percept. Psychophys. **23**, 275–289.

Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. (**1959**). "Multiple observations of signals in noise," J. Acoust. Soc. Am. **31**, 514–521.

Trainer, J. E. (1946). "The auditory acuity of certain birds," thesis/dissertation, Cornell University, Ithaca, NY.

Wiley, R. H., and Richards, D. G. (**1982**). "Adaptations for acoustic communication in birds: sound transmission and signal detection," in *Acoustic Communication in Birds: Communication and Behavior*, edited by D. E. Kroodsma (Academic, New York), pp. 131–181.

Wright, T. F., Cortopassi, K. A., Bradbury, J. W., and Dooling, R. J. (**2003**). "Hearing and vocalizations in the orange-fronted conure (*Aratinga canicularis*)," J. Comp. Psychol. **117**, 87–95.